

**Analyse der Präsenz von *stk33* im Reich  
*Animalia* und Versuche zur Herstellung einer  
Zebrafisch KO-Linie mittels CRISPR/Cas9**

**Dissertation**

Zur Erlangung des Grades

Doktor der Naturwissenschaften

(Dr. rer. nat.)

**Am Fachbereich Biologie der  
Johannes Gutenberg-Universität Mainz**

**Institut für Molekulargenetik,  
gentechnische Sicherheitsforschung und –beratung**

vorgelegt von

Tobias Lautwein

geboren am 16. Juli 1986 in Mainz

Mainz, 2017



Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 14.06.2017





# Inhaltsverzeichnis

Abbildungsverzeichnis.....	IV
Tabellenverzeichnis.....	V
1 Einleitung .....	1
1.1 Proteinkinasen .....	1
1.1.1 Serin/Threonin-spezifische Kinasen.....	2
1.2 Serin/Threonin-Kinase 33.....	3
1.2.1 Die genomische Struktur von <i>stk33</i> .....	3
1.2.2 STK33-Protein.....	5
1.2.3 Lokalisation und Funktion .....	5
1.3 Genome engineering.....	8
1.3.1 Meganukleasen, chemische Nukleasen und Zinkfingernukleasen .....	11
1.3.2 <i>Transcription activator-like effector nucleases</i> (TALENs).....	12
1.3.3 CRISPR/Cas-System .....	14
1.4 Der Zebrafisch ( <i>Danio rerio</i> ) als Modellorganismus.....	17
1.5 Zielsetzung .....	19
2 Material & Methoden .....	20
2.1 Etablierung des Zebrafisches .....	20
2.1.1 Tierhaltung .....	20
2.1.2 Gewinnung von Embryonen.....	21
2.1.3 Präparation adulter Zebrafische .....	21
2.2 Molekularbiologische Methoden zur Analyse von DNA .....	22
2.2.1 Isolierung genomischer DNA.....	22
2.2.2 Polymerase-Kettenreaktion (PCR).....	23
2.2.3 Gelelektrophoretische Auftrennung von DNA.....	24
2.2.4 Wiedergewinnung von DNA-Fragmenten.....	25
2.2.5 Aufreinigung von DNA.....	26
2.2.6 Klonierung von PCR-Produkten.....	26
2.2.7 Sanger-Sequenzierung .....	27
2.2.8 Konzentrations- und Reinheitsbestimmung von DNA .....	28
2.3 Molekularbiologische Methoden zur Analyse von RNA.....	28
2.3.1 Isolierung von Total-RNA .....	28
2.3.2 Konzentrations- und Reinheitsbestimmung von RNA.....	28
2.3.3 Wiedergewinnung von RNA-Fragmenten mittels Elektroelution .....	29
2.3.4 cDNA-Synthese.....	29

2.3.5	<i>in vitro</i> Transkription .....	29
2.3.6	Fällung von RNA .....	30
2.4	genome engineering.....	30
2.4.1	TALENs.....	31
2.4.2	CRISPR/Cas .....	32
2.4.3	Mikroinjektion .....	33
2.4.4	Genotypisierung .....	34
2.5	Bioinformatische Methoden .....	36
2.5.1	Konvertierung der SRA-Datensätze.....	36
2.5.2	Filterung der Datensätze .....	36
2.5.3	<i>De novo</i> Assemblierungen.....	38
2.5.4	Kartierungen.....	38
2.5.5	Genomische Analyse der Klasse <i>Aves</i> .....	38
2.5.6	Analyse des Reichs <i>Animalia</i> mittels „Data-Mining“ .....	39
2.6	Puffer und Lösungen .....	40
3	Ergebnisse.....	42
3.1	Genome engineering.....	42
3.1.1	TALEN-vermittelter „Knock-Out“ von <i>stk33</i> im Zebrafisch .....	42
3.1.2	CRISPR/Cas-vermittelter „Knock-Out“ von <i>no-tail (ntl)</i> im Zebrafisch .....	46
3.1.3	CRISPR/Cas-vermittelter „Knock-Out“ von <i>stk33</i> im Zebrafisch .....	48
3.2	<i>Stk33</i> im Reich <i>Animalia</i> .....	53
3.2.1	<i>Stk33</i> in Neumündern ( <i>Deuterostomia</i> ) .....	53
3.2.2	<i>Stk33</i> in Lophotrochozoen ( <i>Lophotrochozoa</i> ) .....	55
3.2.3	<i>Stk33</i> in Häutungstieren ( <i>Ecdysozoa</i> ) .....	57
3.2.4	<i>Stk33</i> in basalen Tieren .....	62
3.3	<i>Stk33</i> in der Klasse <i>Aves</i> .....	64
3.3.1	<i>Stk33</i> in Urkiefervögeln ( <i>Palaeognathae</i> ) .....	64
3.3.2	<i>Stk33</i> im Haushuhn, <i>Gallus gallus domesticus</i> .....	70
3.3.3	<i>Stk33</i> in Neukiefervögeln ( <i>Neognathae</i> ) .....	71
4	Diskussion.....	75
4.1	Genome Engineering.....	75
4.1.1	TALEN-vermittelte „Knock-Out“ Versuche .....	75
4.1.2	CRISPR/Cas-vermittelte „Knock-Out“ Versuche .....	78
4.2	<i>Stk33</i> im Reich <i>Animalia</i> .....	83
4.3	<i>Stk33</i> in der Klasse <i>Aves</i> .....	90
4.3.1	<i>Stk33</i> in Urkiefervögeln ( <i>Palaeognathae</i> ) .....	91

4.3.2	<i>Stk33</i> in Neukiefervögeln ( <i>Neognathae</i> ).....	92
5	Zusammenfassung .....	99
	Abkürzungsverzeichnis.....	101
	Literaturverzeichnis.....	104
	Anhang .....	116
	A1 Vektorkarten .....	116
	A2 TALEN-vermittelter Knock-out von <i>stk33</i> im Zebrafisch.....	118
	A3 CRISPR/Cas-vermittelter Knock-out von <i>stk33</i> .....	119
	A4 <i>Stk33</i> in der Klasse <i>Aves</i> .....	121
	A5 Übersicht über untersuchte Arten aus dem Reich <i>Animalia</i> .....	124
	A6 Inhaltsverzeichnis des elektronischen Anhangs .....	132
	Curriculum Vitae .....	<b>Fehler! Textmarke nicht definiert.</b>
	Danksagung.....	<b>Fehler! Textmarke nicht definiert.</b>
	Eidesstattliche Erklärung.....	133

# Abbildungsverzeichnis

Abbildung 1.1   Die katalytische Domäne der eukaryotischen Proteinkinasen. ....	2
Abbildung 1.2   Phosphorylierung durch Serin/Threonin-spezifische Kinasen. ....	2
Abbildung 1.3   Der <i>stk33</i> Locus verschiedener Organismen.....	3
Abbildung 1.4   Die genomische Struktur von <i>stk33</i> . ....	4
Abbildung 1.5   Alignment der STK33-Aminosäuresequenz der katalytischen Domäne verschiedener Spezies. ....	6
Abbildung 1.6   Möglichkeiten des Nuklease-induzierten <i>genome engineering</i> .....	9
Abbildung 1.7   Schematische Darstellung der Funktionsweise verschiedener Werkzeuge für das genome engineering. ....	13
Abbildung 1.8   Diversität des CRISPR/Cas-vermittelten adaptiven Immunsystems in Prokaryoten.....	15
Abbildung 1.9   Meilensteine der Zebrafisch Forschung.....	18
Abbildung 2.1   Präparation adulter Zebrafische. ....	21
Abbildung 2.2   verwendete Molekulargewichtsstandards.....	25
Abbildung 2.3   gezielte Mutagenese mittels ZFNs, TALENs und des CRISPR/Cas-Systems.....	31
Abbildung 2.4   Übersicht über die verwendeten CRISPR/Cas-Vektoren.....	33
Abbildung 2.5   Schematische Darstellung der Mikroinjektion.....	34
Abbildung 3.1   TALEN-vermittelter „Knock-Out“ von <i>stk33</i> im Zebrafisch. ....	45
Abbildung 3.2   CRISPR/Cas-vermittelter „Knock-Out“ von <i>no-tail (ntl)</i> im Zebrafisch.....	47
Abbildung 3.3   CRISPR/Cas-vermittelter „Knock-Out“ von <i>stk33</i> im Zebrafisch. (A) .....	49
Abbildung 3.4   Der genomische Locus um <i>stk33</i> in niederen Chordatieren.....	54
Abbildung 3.5   Analyse der aus den identifizierten mRNA-Sequenzen übersetzten Proteinsequenzen mittels InterPro. ....	57
Abbildung 3.6   Alignment ausgewählter, im Reich <i>Animalia</i> gefundener STK33-Proteine. ....	61
Abbildung 3.7   genomische Struktur des <i>stk33</i> Gens im Afrikanischen Vogelstrauß. ....	66
Abbildung 3.8   Analyse des gefundenen Proteins des Vogelstraußes mittels InterPro.....	67
Abbildung 3.9   Aminosäuren Alignment katalytischer Domänen verschiedener Spezies. ....	69
Abbildung 3.10   <i>Stk33</i> -Pseudogen im Haushuhn ( <i>Gallus gallus</i> ). ....	71
Abbildung 3.11   Alignment identifizierter Exons gegen die <i>stk33</i> mRNA Sequenz des Afrikanischen Vogelstrauß. ....	72
Abbildung 4.1   Kladogramm aller untersuchten Taxa aus dem Reich <i>Animalia</i> . ....	87
Abbildung 4.2   Modell einer stabilen genetischen Redundanz.....	96
Abbildung 4.3   Alignment der amino-terminalen Kopfdomäne von Vimentin aus verschiedenen Vogelspezies mit Mensch und Maus.....	98
Abbildung A 1   Übersicht über die verwendeten Vektoren. ....	116
Abbildung A 2   Übersicht über die verwendeten Vektoren. ....	117
Abbildung A 3   Elektrophoretische Analyse des T7EI-Assay des TALII-Vektors. ....	118
Abbildung A 4   Chromtagramm der Sequenzierung der assemblierten TALEN-Vektoren.....	118
Abbildung A 5   Gelelektrophoretische Analyse des T7EI-Assay der <i>stk33</i> -I „off-target“ Zielsequenz.....	119
Abbildung A 6   Elektrophoretische Analyse des T7EI-Assay der adulten F <sub>0</sub> -Generation. ....	119

Abbildung A 7   Analyse des für die Genotypisierung verwendeten Bereiches hinsichtlich möglicher Varianz. ....	120
Abbildung A 8   Phylogenetischer Stammbaum der Klasse <i>Aves</i> mit Kennzeichnung untersuchter Familien. ....	121
Abbildung A 9   Alignment identifizierter <i>stk33</i> Exons aus den verbleibenden Vogelspezies gegen die <i>stk33</i> mRNA Sequenz des Vogelstraußes. ....	122
Abbildung A 10   STK33 Phylogenie. ....	123
Abbildung A 11   Stammbaum der Ordnung der Hautflügler ( <i>Hymenoptera</i> ). ....	131

## Tabellenverzeichnis

Tabelle 2.1   PCR Ansatz Q5® DNA Polymerase .....	23
Tabelle 2.2   PCR Ansatz GoTaq® DNA Polymerase.....	23
Tabelle 2.3   PCR-Programm .....	23
Tabelle 2.4   „long range“ PCR-Ansatz .....	24
Tabelle 2.5   „long-range“ PCR-Programm.....	24
Tabelle 2.6   Ligationsansatz .....	27
Tabelle 2.7   Ansatz für die <i>in vitro</i> Transkription der sgRNA .....	30
Tabelle 2.8   Programm zur De- und Renaturierung. ....	35
Tabelle 2.9   Übersicht über alle für die Analyse der Klasse <i>Aves</i> verwendeten Next-Generation Sequencing Datensätze. ....	37
Tabelle 3.1   Übersicht über die für die Zielsequenz <i>stk33</i> -I durchgeführten Mikroinjektionsrunden. ....	50
Tabelle 3.2   Übersicht über untersuchte Arten innerhalb der <i>Deuterostomia</i> und <i>Lophotrochozoa</i> .....	56
Tabelle 3.3   Übersicht über untersuchte Arten innerhalb der <i>Ecdysozoa</i> . ....	58
Tabelle 3.4   Übersicht über untersuchte Arten innerhalb der Klasse <i>Insecta</i> . ....	62
Tabelle 3.5   Übersicht über untersuchte Arten innerhalb der niederen Tiere. ....	63
Tabelle 3.6   Übersicht über untersuchte Arten innerhalb der Klasse <i>Aves</i> . ....	74
Tabelle 4.1   Bekannte Vimentin-phosphorylierende Kinasen und deren Phosphorylierungsstellen.....	98
Tabelle A 1   Detaillierte Übersicht über untersuchte Arten innerhalb der <i>Insecta</i> . ....	124



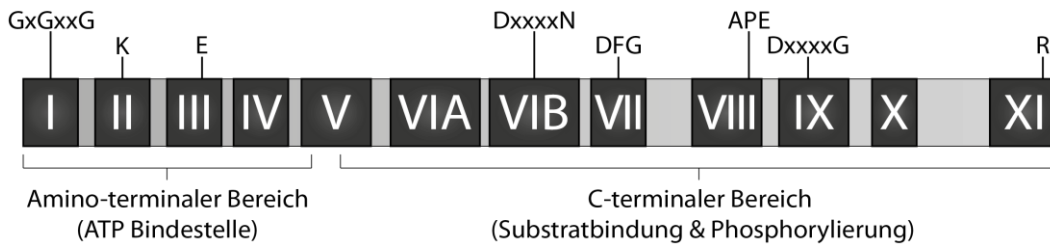
# 1 | Einleitung

## 1.1 Proteinkinasen

Proteinkinasen sind Enzyme, die in der Lage sind, endständige  $\gamma$ -Phosphatgruppen von einem Nucleosidtriphosphat (z.B. ATP) reversibel auf freie Hydroxylgruppen der Aminosäureseitenketten des Substrates zu übertragen [1; 2]. Durch die Übertragung des negativ geladenen Phosphates (Abbildung 1.2) kommt es im Substratprotein zu veränderten elektrostatischen Wechselwirkungen und zur Ausbildung von Wasserstoffbrücken, was zu einer Konformationsänderung und damit zur Aktivierung bzw. Inaktivierung des Substrates führen kann [2]. Die Phosphorylierung durch Proteinkinasen spielt, in Wechselwirkung mit der Dephosphorylierung durch Phosphatasen [2], eine zentrale Rolle in nahezu allen zellulären Prozessen. Proteinkinasen kontrollieren und regulieren Metabolismus, Transkription, Zellteilung und -bewegung, programmierten Zelltod und partizipieren bei der Immunantwort und an der Funktion des Nervensystems [3; 4]. Daher sind Mutationen von Kinasen oft mit Krebserkrankungen assoziiert [2]. Das menschliche Kinom besteht aus über 500 verschiedenen Kinasen, die in ihrer Gesamtheit etwa 30 % aller zellulären Proteine phosphorylieren können, jedoch nur etwa 2 % bis 3 % der zellulären Proteine ausmachen [3; 4].

Die Klassifizierung von Kinasen erfolgt aufgrund ihrer Akzeptor-Aminosäure, auf die die Phosphatgruppe übertragen wird. Auf diese Weise wird zwischen den vier verschiedenen Klassen unterschieden, den Serin/Threonin-spezifischen Kinasen (vgl. 1.1.1), Tyrosin-spezifischen Kinasen, Histidin-spezifischen Kinasen und Aspartat- bzw. Glutamat-spezifischen Kinasen [5].

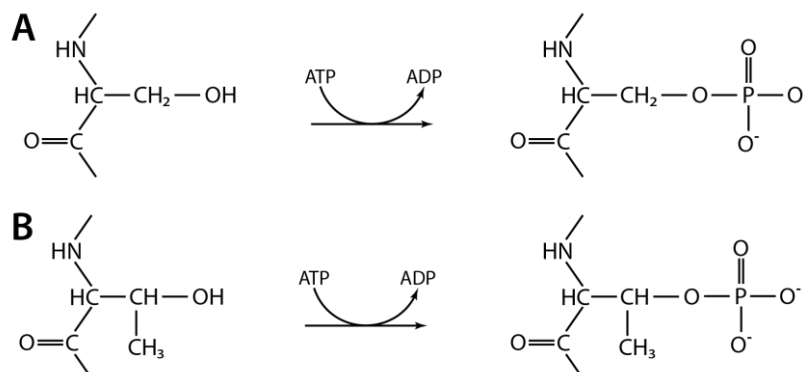
Eine ca. 250 Aminosäuren lange Kinasedomäne ist typisch für die Familie der eukaryotischen Proteinkinasen, eine der größten Proteinsuperfamilien [6]. Diese Kinasedomäne besteht aus 12 konservierten Subdomänen und lässt sich in einen amino-terminalen Bereich, welcher die ATP Bindestelle enthält, und einen C-terminalen Bereich, der für die Substratbindung und den Transfer der Phosphatgruppe auf das Substrat zuständig ist, unterteilen (Abbildung 1.1) [6; 7].



**Abbildung 1.1 | Die katalytische Domäne der eukaryotischen Proteinkinase.** Subdomänen sind durch römische Ziffern gekennzeichnet, hoch konservierte Motive sind darüber dargestellt. Die Domäne eukaryotischer Proteinkinase besteht aus dem amino-terminalen Bereich, der die ATP-Bindestelle enthält, und dem C-terminalen Bereich, welcher für Substratbindung und Phosphorylierung zuständig ist. Für die Bindung und Ausrichtung des ATPs sind v.a. das Lysin der Subdomäne II sowie das Aspartat der Subdomäne VII von Bedeutung. Bei dem konservierten Aspartat innerhalb der Subdomäne VI handelt es sich höchst wahrscheinlich um die katalytische Aminosäure für den Transfer des Phosphatrestes auf das Substrat. (verändert nach: Hanks, SK: *Genomic analysis of the eukaryotic protein kinase superfamily: a perspective*. Genome Biol. 2003, 5:111)

### 1.1.1 Serin/Threonin-spezifische Kinasen

Serin/Threonin-spezifische Kinasen katalysieren die Übertragung von  $\gamma$ -Phosphatresten von ATP (bzw. GTP) über eine Veresterung auf den Alkoholrest der Aminosäuren Serin und Threonin (Abbildung 1.2). Innerhalb des Substrates werden nur bestimmte Serine und Threonine phosphoryliert, welche durch die flankierenden Sequenzen und die sterische Zugänglichkeit bestimmt werden [5]. Bei etwa 25 % aller Proteinkinase handelt es sich um Serin/Threonin-spezifische Kinasen [8]. Aufgrund der Beteiligung an nahezu allen zellulären Prozessen [3; 4] ist eine strikte Regulierung der Kinasen notwendig. Diese erfolgt v.a. durch Aktivatoren bzw. Inhibitoren oder durch Phosphorylierung bzw. Dephosphorylierung der Kinasen [5]. Mit Bezug auf unterschiedliche Substratspezifitäten, notwendige Ko-Faktoren oder die Empfindlichkeit gegenüber Inhibitoren kann die Familie der Serin/Threonin-spezifischen Kinasen in weitere Unterfamilien eingeteilt werden [5].



**Abbildung 1.2 | Phosphorylierung durch Serin/Threonin-spezifische Kinasen.** Diese sind in der Lage die  $\gamma$ -Phosphatreste von Nucleosidtriphosphaten (ATP bzw. GTP) über eine Veresterung auf den Alkoholrest der Aminosäuren (A) Serin und (B) Threonin zu übertragen. (verändert nach Krauss, Gerhard (2003): *Biochemistry of signal transduction and regulation*. 3. Aufl. Weinheim: Wiley-VCH)

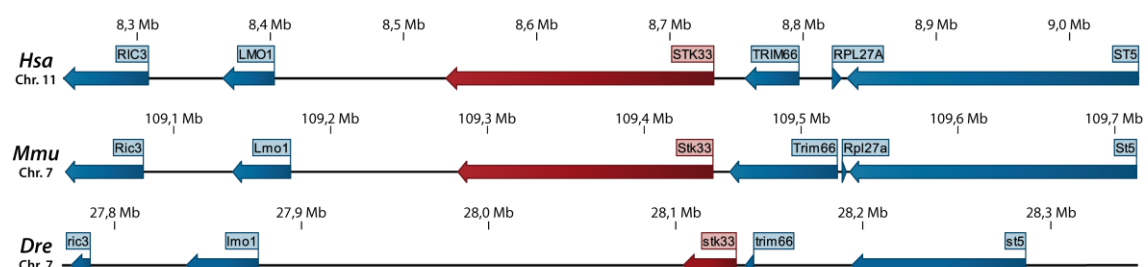


Die Serin/Threonin-spezifische Kinase 33 gehört hierbei zu der Subfamilie der Calcium/Calmodulin-abhängigen Kinasen (CAMK) [9], die durch die An- bzw. Abwesenheit von Calcium ( $\text{Ca}^{2+}$ ) reguliert werden [5; 10]. Bei Abwesenheit von Calcium ist das aktive Zentrum der CAMKs durch die benachbarte inhibitorische Domäne blockiert. Ein steigender Calcium-Spiegel sorgt dann für eine Absättigung der Kinase mit vier  $\text{Ca}^{2+}$ -Ionen, was zu einer Konformationsänderung und somit zur Exposition des aktiven Zentrums führt [10]. Dies ermöglicht eine Autophosphorylierung der Kinasen, welche eine weitere Ebene der Regulation darstellt [11; 12]. Innerhalb der Subfamilie der Calcium/Calmodulin-abhängigen Kinasen finden sich sowohl Kinasen mit sehr spezifischer Substratspezifität („*Myosin light chain kinase*“), als auch solche mit sehr breiter Substratspezifität (CAMKII) [5; 10].

## 1.2 Serin/Threonin-Kinase 33

Im Rahmen des Humangenomprojektes [13] wurden die Chromosomenregion 11p15.3 des Menschen und die homologe Region in der Maus auf Chromosom 7 sequenziert [14; 15]. Hierbei handelt es sich um eine genreiche Region mit klinischer Relevanz, da diese Region mit verschiedenen Krankheitsbildern wie dem Beckwith-Wiedemann-Syndrom, Hämoglobinopathie, Diabetes Typ 1 und T-Zell-Leukämie [16], sowie verschiedenen Krebsarten in Verbindung gebracht wird [17]. Durch eine komparative Genomanalyse dieser Region in Mensch und Maus konnte die Serin/Threonin-Kinase 33 (*stk33*) in dieser syntänen Region (vgl. Abbildung 1.3) als neues Gen identifiziert werden [9].

### 1.2.1 Die genomische Struktur von *stk33*

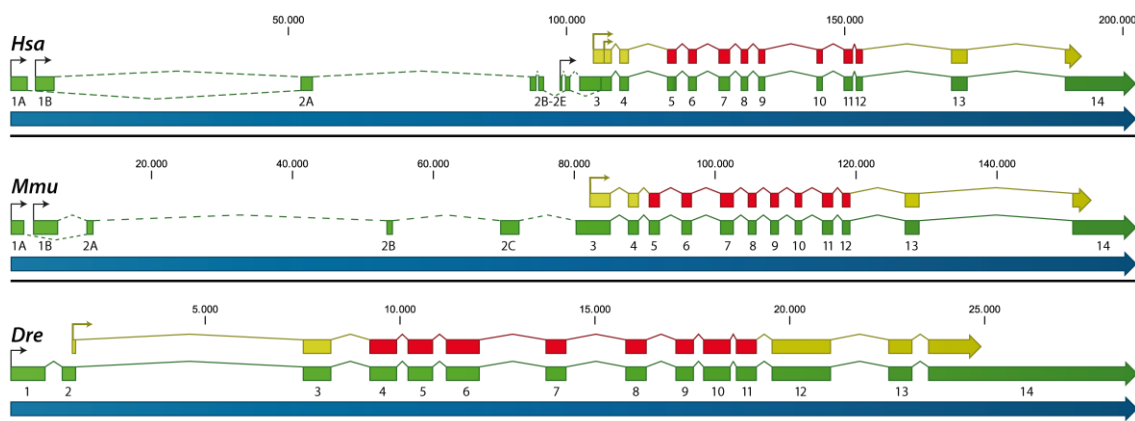


**Abbildung 1.3 | Der *stk33* Locus verschiedener Organismen.** Die genomische Region um den Locus von *stk33* weist innerhalb der *Vertebrata* eine hohe Syntänie in dem Bereich zwischen den Genen *st5* und *ric3* auf. Das im Zebrafisch fehlende Gen *rpl27a* findet sich hier auf Chromosom 25.

Wie die Abbildung 1.3 zeigt, ist die Chromosomenregion, in der *stk33* liegt, innerhalb der Vertebraten hoch konserviert. Das *stk33*-Gen liegt zwischen den Genen „*Suppression Of Tumorigenicity 5*“ (*st5*) und „*Resistant To Inhibitor Of Cholinesterase 3*“ (*ric3*). Im Zebrafisch liegt diese syntäne Region ebenfalls auf Chromosom 7, wobei sich hier das Gen für das Ribosomale Protein L27a (*rpl27a*) auf Chromosom 25 befindet.

*Stk33* besteht aus 14 Exons, wobei in Mensch und Maus die Exons 5 bis 12 für die Kinasedomäne kodieren, im Zebrafisch wird diese durch die Exons 4 bis 11 kodiert [18]. Das aktive Zentrum wird im Menschen durch Exon 8 und im Zebrafisch durch Exon 7 kodiert. Im Menschen finden sich drei verschiedene Transkriptionsstartpunkte, während sich in der Maus nur zwei finden [19]. Die 5′ untranslatierte Region (5′-UTR) zeichnet sich durch eine hohe Variabilität im Bereich des zweiten Exons aus. So können im Menschen fünf (2A-2E) und in der Maus drei verschiedene Exons (2A-2C) in unterschiedlichen Transkripten als Exon 2 fungieren [18; 19]. Im Mensch findet sich zudem eine zweite Translationsstartstelle im Exon 3 (siehe Abbildung 1.4) [19].

Innerhalb der *Vertebrata* konnte das Vorhandensein von *stk33* bereits in den Gruppen der Amphibien (*Amphibia*), Knochen- und Knorpelfische (*Osteichthyes* und *Chondrichthyes*), Säugetiere (*Mammalia*) und der Reptilien (*Reptilia*) nachgewiesen werden (vgl. Abbildung A 10). Innerhalb der Klasse der Vögel (*Aves*) und außerhalb der *Vertebrata* konnten hingegen bisher keine Erkenntnisse über die Präsenz von *stk33* gewonnen werden.



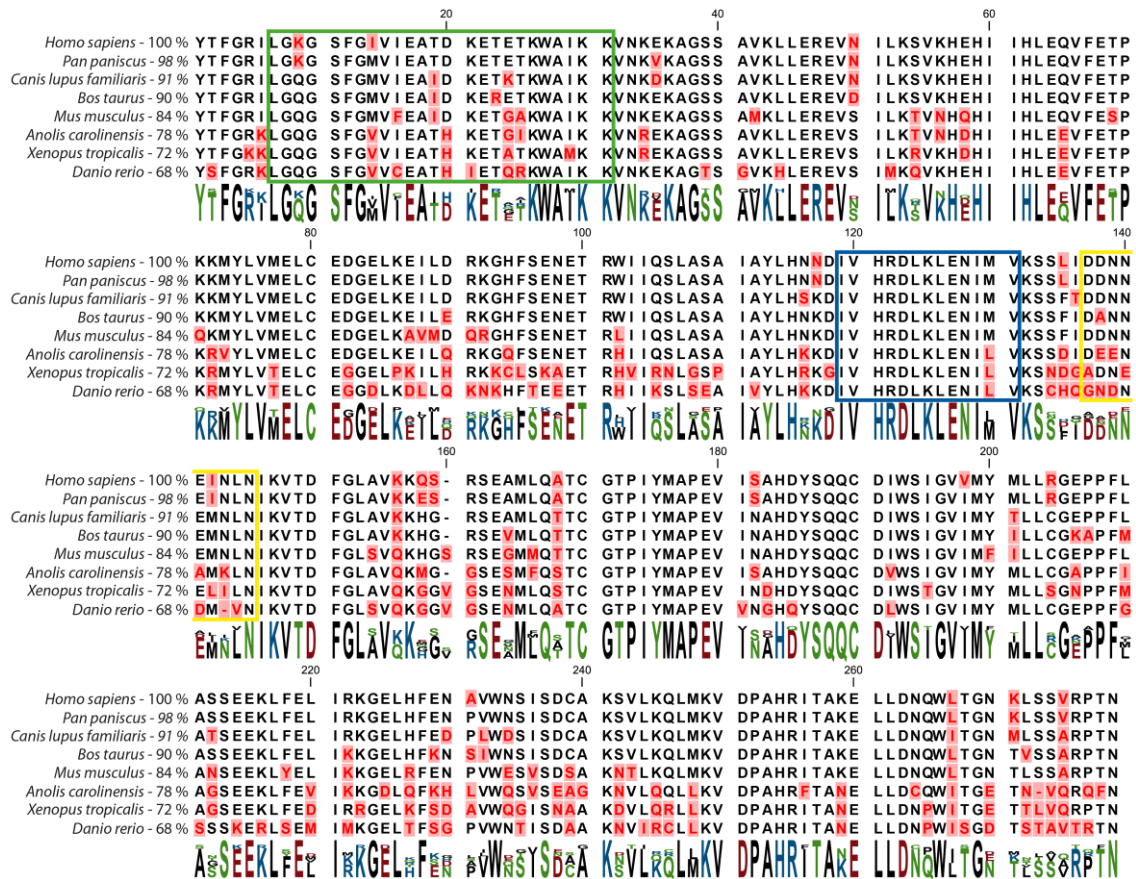
**Abbildung 1.4 | Die genomische Struktur von *stk33*.** *Stk33* besteht in Mensch (*Hsa*), Maus (*Mmu*) und Zebrafisch (*Dre*) aus 14 Exons, wobei bei Mensch und Maus die Exons 2A bis 2E in unterschiedlichen Transkripten vorkommen (durch gestrichelte Linien gekennzeichnet). Genannotationen sind in Blau, mRNA in Grün und der kodierende Bereich ist in Gelb dargestellt. Exons, welche für die Kinase-Domäne kodieren, sind rot hervorgehoben. Transkriptionsstartpunkte sind durch schwarze Pfeile, Translationsstartpunkte durch gelbe Pfeile gekennzeichnet. Alle Introns sind mit 5 % ihrer tatsächlichen Größe dargestellt.

### 1.2.2 STK33-Protein

Der offene Leserahmen (ORF) von *stk33* beläuft sich im Menschen auf 1545 bp und kodiert für ein Protein mit einem Molekulargewicht von 57,8 kDa. Der murine ORF ist 1417 bp lang und kodiert für ein 54,4 kDa großes Protein [9]. Die STK33-Proteine zeigen eine sehr starke Konservierung der katalytischen Kinasedomäne (84 % innerhalb der Kinasedomäne) [19] (vgl. Abbildung 1.5). Durch phylogenetische Analysen konnte STK33 der Subfamilie der Calcium/Calmodulin-abhängigen Kinasen (1.1.1) zugeordnet werden, obwohl weder die typische Calcium/Calmodulin Bindedomäne noch die C-terminale regulatorische Domäne vorhanden sind [9]. Aufgrund der Sequenzähnlichkeit und der Tatsache, dass auch andere Mitglieder der Calcium/Calmodulin-abhängigen Kinasen nicht über eine Calcium/Calmodulin Bindedomäne verfügen, wird STK33 jedoch trotzdem dieser Subfamilie zugeordnet [7]. Innerhalb der Kinasedomäne finden sich sowohl eine ATP-Bindestelle, als auch das aktive Zentrum (vgl. Abbildung 1.5), sodass alle Bestandteile einer aktiven Kinase vorhanden sind [9]. Beim Vergleich von STK33 mit anderen Kinasen zeigte sich, dass sich in unmittelbarer Nachbarschaft zum aktiven Zentrum eine Asx-reiche Schleife befindet (Abbildung 1.5), welche als spezifisch für STK33 beschrieben wurde [19].

### 1.2.3 Lokalisation und Funktion

Erste Untersuchungen zur Lokalisation in humanen und murinen Geweben zeigten, dass *stk33* in den meisten Geweben nur schwach exprimiert ist und das Expressionsmuster dem anderer CAMKs ähnelt [20]. Während der Embryonalentwicklung konnte eine Expression in verschiedenen embryonalen Geweben wie Herz, Gehirn, Lunge und Rückenmark nachgewiesen werden. In adulten Geweben ist *stk33* im Lungenepithel, in alveolaren Makrophagen und in Horizontalzellen der Retina exprimiert. Die höchste Expression von *stk33* konnte jedoch in Testes, insbesondere in Spermatozyten, nachgewiesen werden [20]. Bei der Untersuchung von Proteinextrakten aus Testes junger Mäuse (10 dpp – „days postpartum“) und älterer Mäuse (20 dpp; 30 dpp) konnte STK33 nur in den Extrakten der älteren Mäuse während einer Entwicklungsphase, in der die Spermatogenese abläuft [21], nachgewiesen werden [20]. Diese Ergebnisse und die lokale Expression in einigen embryonalen Geweben führten zu der Hypothese, dass STK33 möglicherweise an Organogenese und Spermatogenese beteiligt ist [20]. Eine mögliche Beteiligung an der Organogenese wird durch die Beobachtung gestärkt, dass *stk33* zu den am stärksten herunterregulierten Genen bei Patienten mit Fallot-Tetralogie gehört, welche zu verschiedenen Fehlbildungen des Herzens führt [22].



**Abbildung 1.5 | Alignment der STK33-Aminosäuresequenz der katalytischen Domäne verschiedener Spezies.** Die prozentuale Sequenzidentität innerhalb der katalytischen Domäne ist neben den jeweiligen Sequenzen dargestellt. Die ATP-Bindestelle ist durch ein grünes Rechteck, das aktive Zentrum durch ein blaues Rechteck und die Asx-reiche Schleife durch ein gelbes Rechteck markiert. Nicht-konservierte Aminosäuren sind in Rot hervorgehoben.

Immunfluoreszenz-Versuche zeigten eine deutliche Ko-Lokalisation von STK33 mit dem Typ-III Intermediärfilament Vimentin in verschiedenen Regionen des Gehirns von Maus, Ratte und Hamster [23]. Hilfe eines Kinase-Assays konnte nachfolgend gezeigt werden, dass STK33 *in vitro* in der Lage ist die amino-terminale Kopfdomäne von Vimentin zu phosphorylieren, Außerdem zeigten diese Versuche, dass die Autophosphorylierung von STK33 eine Voraussetzung für dessen Aktivität ist [24]. Eine unmittelbare Assoziation von STK33 und Vimentin *in vivo* konnte durch Ko-Immünpräzipitationsversuche und Ko-Sedimentationsversuche nachgewiesen werden [24]. Diese Ergebnisse führten zu der Vermutung, dass STK33 über die spezifische Phosphorylierung von Vimentin in die Dynamik und Regulierung des Auf- und Abbaus des Intermediärfilamentes involviert ist [23; 24]. Ergebnisse einer weiteren Studie deuten zudem daraufhin, dass STK33 auch eine Komponente des neuroendokrinen Netzwerkes sein könnte [25].

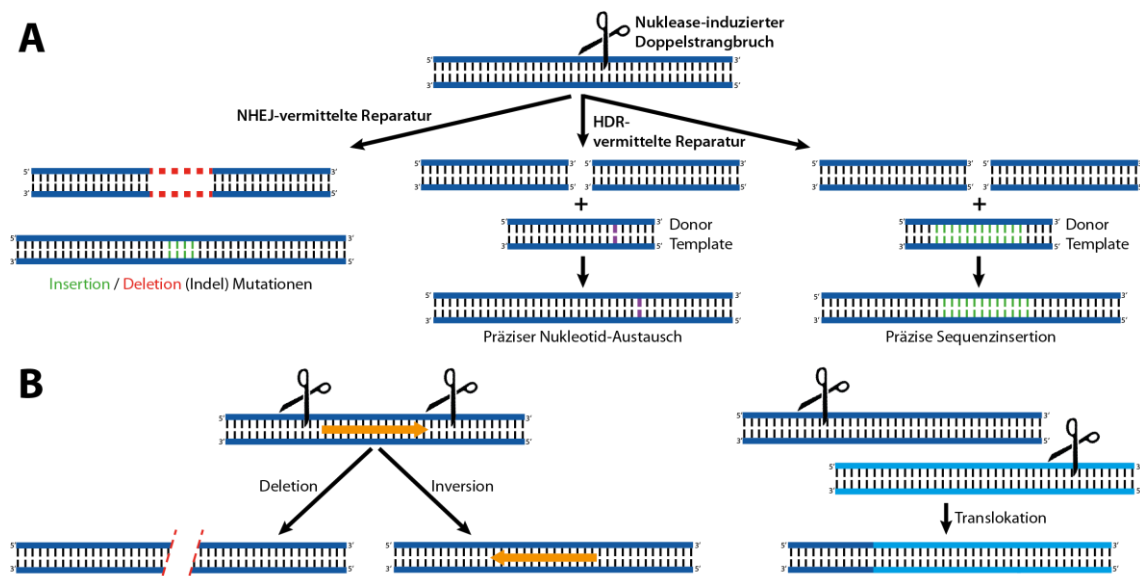
Des Weiteren wurde STK33 mittels Hochdurchsatz RNA-Interferenz (RNAi) Versuchen in KRAS-mutanten humanen Krebszellen als essenziell für deren Überleben identifiziert [26–28], ohne selbst als Onkogen zu fungieren („*non-oncogene addiction*“) [29]. Jedoch konnte eine spätere

Studie von Babij et al. diese Ergebnisse mit Hilfe von RNAi- und Überexpressionsversuchen sowie mit Inhibitoren für STK33 nicht bestätigen [30]. Allerdings wurden öffentliche Zweifel an den Ergebnissen von Babij *et al.* laut, welche die verwendete Methodik in Frage stellten [26]. Doch auch nachfolgende Studien konnten unter Verwendung von STK33-Inhibitoren keinen Effekt auf die Viabilität von KRAS-abhängigen Krebszellen feststellen [31–33]. Trotz alledem konnten mehrere folgende Studien einen Zusammenhang zwischen STK33 und Dickdarmkrebs [34], dem Leberzellkarzinom [35], dem Plattenepithelkarzinom [36], dem Lungenkrebs [37] und dem Epstein-Barr Virus (EBV) positiven Burkitt-Lymphom [38] herstellen, sodass eine Rolle von *stk33* in verschiedenen Krebsarten plausibel erscheint.

## 1.3 Genome engineering

Der Begriff des „*Genome engineering*“ umfasst alle Techniken zur zielgerichteten Manipulation des Genoms lebender Organismen. Die erste Methode zur Ausschaltung eines Gens wurde von Mario R. Capecchi, Oliver Smithies und Martin J. Evans in Mäusen während der frühen 1980er Jahre etabliert und mit dem Nobelpreis für Physiologie und Medizin 2007 honoriert [39]. Da diese Methode aber ausschließlich auf homologer Rekombination basiert, ist deren Effizienz in Zellen höherer Eukaryoten mit Mutationsraten von  $10^{-6}$  bis  $10^{-7}$  entsprechend niedrig [40]. Auch Transposons und Rekombinasen und wurden für Genommodifikationen eingesetzt, sind jedoch ebenfalls aufwendig in der Anwendung und vergleichsweise ineffizient [41; 42].

Daher wandte man sich in den folgenden Jahren den sogenannten molekularen Scheren („*molecular scissors approach*“) zu. Hierbei werden durch verschiedene Nukleasen Doppelstrangbrüche an der Ziel-Sequenz eingeführt, was zur Aktivierung des NHEJ-Reparaturprozesses („*non-homologous end joining*“) und des HDR-Reparaturprozesses („*homology directed repair*“) durch die Zelle führt. Dies ermöglicht eine Vielzahl an verschiedenen, zielgerichteten Manipulationen des Genoms (vgl. Abbildung 1.6). Während bei dem HDR-Reparaturprozess das homologe Chromosom bzw. Chromatid als Vorlage dient und so Fehler bei der Reparatur verhindert, handelt es sich bei der NHEJ-vermittelten Reparatur um ein Matritze-unabhängiges Verfahren, wodurch die Reparatur des Doppelstrangbruches fehlerbehaftet ist. Hierbei werden sowohl kleinere Insertionen als auch Deletionen (Indels) eingefügt [43]. Geschieht dies innerhalb des kodierenden Bereiches eines Gens, so kann hierdurch der offene Leserahmen zerstört werden, was zu einem Ausschalten („*Knock-Out*“) dieses Gens führt. Mit Hilfe der HDR-vermittelten Reparatur können bei Ko-Injektion eines geeigneten, homologen Donor DNA-Moleküls präzise Nukleotidaustausche durchgeführt werden. Des Weiteren können zielgerichtete Insertionen bestimmter Nukleotidsequenzen durchgeführt werden, welche komplette Gene oder andere genetische Elemente beinhalten können [44]. Bei der Verwendung von zwei verschiedenen Nukleasen können größere genomische Bereiche deletiert oder invertiert werden, sofern die beiden Nukleasen auf demselben Chromosom schneiden [45; 46]. Schneiden diese jedoch auf unterschiedlichen Chromosomen, sind Translokationen zwischen diesen Chromosomen möglich (vgl. Abbildung 1.6) [47]. Dieses Verfahren ermöglicht eine zielgerichtete Korrektur von solchen genetischen Defekten. Aufgrund der vielseitigen Möglichkeiten genetischer Manipulationen kommen diese Methoden in biomedizinischen Forschung, Biotechnologie und der Genterapie zum Einsatz.



**Abbildung 1.6 | Möglichkeiten des Nuklease-induzierten genome engineering.** (A) Ein durch eine Nuklease induzierter Doppelstrangbruch (DSB) kann entweder durch die NHEJ-vermittelte Reparatur („non-homologous end joining“) oder durch die HDR-vermittelte Reparatur („homology directed repair“) beseitigt werden. Bei der NHEJ-vermittelten Reparatur kommt es zur Einführung von Indels an der Stelle des DSB. Eine Reparatur mittels homologer Rekombination in Verbindung mit einer geeigneten Donor-DNA kann zu einem präzisen Austausch von Nukleotiden oder zur Insertion bestimmter Sequenzabschnitte führen. (B) Die Einführung zweier DSB durch verschiedene Nukleasen auf dem gleichen Chromosom kann zur Deletion oder Inversion des dazwischenliegenden DNA-Abschnittes führen. Schneiden beide Nukleasen auf unterschiedlichen Chromosomen, so kann dies zu einer Translokation führen. (verändert nach: Joung, JK; Sander, JD: TALENs: a widely applicable technology for targeted genome editing. Nat. Rev. Mol. Cell Biol. 2013, 1:49–55)

In der biomedizinischen Forschung werden diese Techniken für die reverse Genetik hauptsächlich zwecks Funktionsanalysen von Genen, aber auch zur Analyse von nicht-kodierenden Elementen, wie microRNAs (miRNAs) [48; 49] und long-non coding RNAs (lncRNAs) [50], eingesetzt. Eine weitere Anwendungsmöglichkeit ist die funktionelle Analyse von Einzelnukleotid-Polymorphismen (SNPs) durch deren gezielte Deletion oder Insertion in einem ansonsten isogenen Hintergrund [44; 51; 52]. Ein großer Vorteil dieser Technologie ist die Tatsache, dass diese Nukleasen direkt in frühe Embryonen injiziert werden können, was die Verwendung von embryonalen Stammzellen überflüssig macht [53–57] und auch die Herstellung von Krankheitsmodellen in verschiedenen Tieren deutlich erleichtert [54; 58]. Im Bereich der Biotechnologie finden diese Nukleasen Anwendung bei der Herstellung gentechnisch veränderter Pflanzen [59–61] und Nutztiere [62–64]. Ein weiteres Anwendungsgebiet ist die umstrittene Gentherapie am Menschen. Hierunter versteht man die Veränderung des menschlichen Genoms mit dem Ziel, genetisch bedingte Krankheiten heilen zu können. Der große Vorteil dieser Methode liegt darin, dass damit die Ursache einer Krankheit und nicht nur deren Symptome behandelt werden. Hierzu werden dem Patienten Zellen des betroffenen Organs entnommen, *in vitro* der Behandlung unterzogen und dann wieder in den Patienten eingebracht. Sollte dies nicht möglich sein, kann die Gentherapie auch

*in vivo* erfolgen [65]. Jedoch treten bei solchen Gentherapien immer wieder Probleme und z.T. gravierende Nebenwirkungen auf. So müssen die DNA-Moleküle in die Zellen eingebracht und dort aktiviert werden. Hierzu werden hauptsächlich inaktivierte Adenoviren verwendet [65], was häufig zu heftigen Immunreaktionen führt, welche den Tod des Patienten zur Folge haben können [28]. Des Weiteren kann das vermeintlich heilende Gen in andere essenzielle Gene des Patienten springen und diese zerstören. Dies geschah bei einer klinischen Studie zur Behandlung des „*severe combined immunodeficiency syndrome*“ (SCID), woraufhin ein Teil der Patienten Leukämie entwickelte [66; 67]. Bei der Genommodifikation mittels programmierbarer Nukleasen kann dieses Problem umgangen werden, da mit Hilfe dieser Nukleasen die defekten Gene des Patienten „repariert“ werden können und nicht mehr ersetzt werden müssen. Aber auch hier muss die Möglichkeit einer unspezifischen Genommodifikation („*off target effects*“) berücksichtigt werden [68]. Weiterhin wirft die Gentherapie eine Reihe an ethischen Fragen auf, insbesondere nachdem es gelungen ist, mittels des CRISPR/Cas-Systems die Genome von Primaten [69] und sogar von menschlichen Embryonen [70] zu modifizieren. Aufgrund dessen haben bisher lediglich zwei Gentherapeutika weltweit eine Zulassung erhalten. Das erste zugelassene Gentherapeutikum wurde 2003 unter dem Namen „Gendicine“ in China zugelassen. Hierbei handelt es sich um einen nicht replikationsfähigen Adenovirus-Vektor, mit dem das defekte Tumorsuppressorgen p53 in Kopf-Hals-Plattenepithelkarzinomen ersetzt werden soll [71]. 2012 wurde das zweite Medikament, welches unter dem Namen „Glybera“ vermarktet wird, in der EU zugelassen. Dieses Medikament nutzt ebenfalls einen Adenovirus (AAV1) als Vektor und soll das defekte Gen für die Lipoproteinlipase in den Muskelzellen der Patienten mit der extrem seltenen Lipoproteinlipase-Defizienz durch eine funktionale Kopie ersetzen [72]. Mit Behandlungskosten von 1,1 Millionen Euro pro Patient ist Glybera das z.Z. teuerste Medikament am Markt [73].

*Ex vivo* konnten bereits erfolgreich Korrekturen defekter Gene in vom Patienten entnommenen Zellen vorgenommen werden, die an Krankheiten wie Epidermolysis bullosa [74], Muskeldystrophie Duchenne [75], Mukoviszidose [76], Sichelzellenanämie [51; 77], Alpha-1-Antitrypsin-Mangel [78] und Septischer Granulomatose [79] litten. Insbesondere für die Behandlung des „*acquired immunodeficiency syndrome*“ (AIDS) birgt die Gentherapie große Hoffnungen, nachdem es gelang, einen Patienten durch die Knochenmarktransplantation eines Spenders mit einem mutanten CCR5-Rezeptor zu heilen [80]. Jedoch ist die Suche nach kompatiblen Spendern, die zusätzlich ein mutiertes CCR5-Rezeptorgen besitzen, nur selten erfolgreich. Ein vielversprechender Ansatz ist hier die Entnahme von Knochenmark, das Ausschalten des CCR5-Gens *in vitro* und die Rückgabe der Zellen in den Patienten [81].



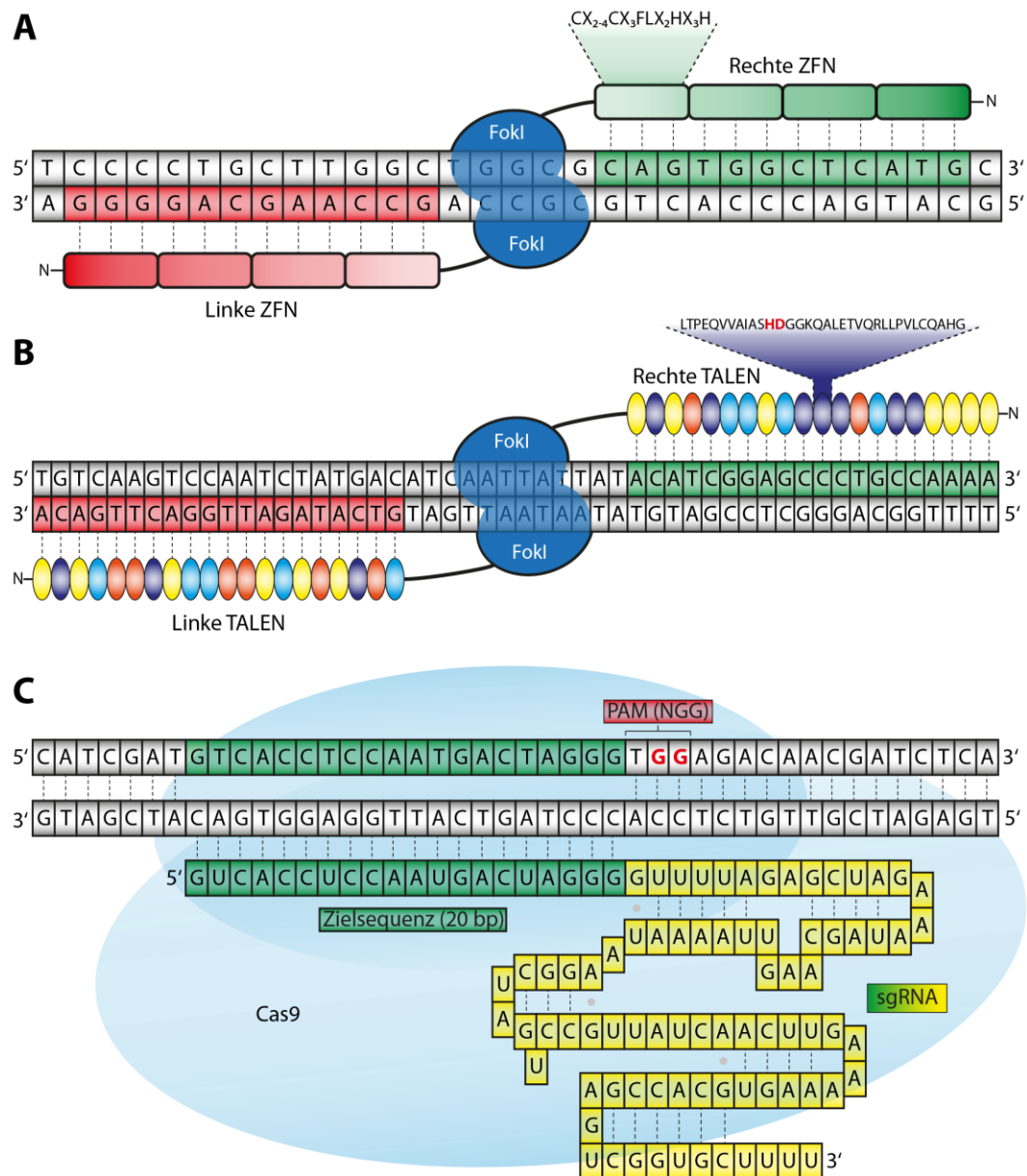
### 1.3.1 Meganukleasen, chemische Nukleasen und Zinkfingernukleasen

Mit der Verwendung von Meganukleasen Anfang der 1990er Jahre gab es erstmals die Möglichkeit gezielt Doppelstrangbrüche in komplexe Genome einführen zu können. Bei Meganukleasen handelt es sich um natürlich vorkommende, sequenzspezifische Endonukleasen mit einer Erkennungssequenz von 14 bp bis 40 bp [82]. Aufgrund der nicht vorhandenen Modularität der DNA-Bindedomäne ist die Herstellung individueller Meganukleasen extrem aufwendig und zeitintensiv, da sowohl die Erkennung und Bindung der DNA, als auch die Nuklease-Aktivität in einer einzelnen Domäne verflochten sind [82–84]. Ein weiteres Werkzeug zur gezielten Modifikation von Genomen sind chemische Endonukleasen. Diese bestehen aus einer chemischen DNA-Bindedomäne, wie „*triplex-forming oligonucleotides*“ (TFO) oder Polyamin und einem chemischen oder enzymatischen Effektor-Molekül. Der Vorteil dieser Methode liegt darin, dass die Sequenzspezifität *a priori* festgelegt werden kann und der aufwändige Mutationsprozess sowie das anschließende Screening nach neuen DNA-Bindesepezifitäten entfallen. Jedoch bestehen nach wie vor Probleme beim Einbringen der Moleküle in die Zelle und mit deren Effizienz [83; 85].

Die Zinkfingernukleasen (ZFNs) gehören zusammen mit den TALENs (1.3.2) und dem CRISPR/Cas-System (1.3.3) zu den sogenannten programmierbaren Nukleasen. Durch die Verwendung von ZFNs wurde es erstmals möglich, mit einem überschaubaren Aufwand gezielte Genommodifikationen durchzuführen. In den folgenden Jahren wurden diese erfolgreich bei einer Vielzahl an Organismen wie Viren, Bakterien, Pflanzen und verschiedenen Tieren, einschließlich Säugetieren, verwendet [86–89]. Eine Zinkfingernuklease besitzt eine modulare Struktur, welche aus drei bis vier Zinkfingerproteinen (ZFPs) besteht, wobei jedes dieser ZFPs ein Nukleotid-Triplett bindet, und aus einer *FokI*-Nuklease Domäne [90]. Um mittels ZFNs einen Doppelstrangbruch erzeugen zu können, müssen zwei ZFNs verwendet werden, da eine Dimerisierung zweier *FokI*-Domänen für deren enzymatische Aktivität notwendig ist (vgl. Abbildung 1.7 A) [91]. Zwar kann durch eine modulare Kombination verschiedener ZFPs die ZFN mit der gewünschten Spezifität erstellt werden, jedoch zeigte sich, dass solche ZFNs die DNA oft nur ineffizient schneiden [92] oder toxisch sind [93]. Erst durch Zell-basierte Selektionsmethoden und die Berücksichtigung der Abhängigkeiten zwischen benachbarten ZFPs konnten effektive ZFNs hergestellt werden [94–96]. Dies macht die Herstellung solcher ZFNs im Vergleich zu TALENs (1.3.2) und CRISPR/Cas9 (1.3.3) jedoch zeitaufwendig und kostenintensiv. Außerdem gibt es keine öffentlich zugängliche Sammlung aller 64 Zinkfinger, die notwendig wäre, um jede mögliche Sequenz abzudecken. Dadurch ergeben sich zusätzlich Einschränkungen bei der Auswahl der Zielsequenzen [97; 98].

### 1.3.2 *Transcription activator-like effector nucleases (TALENs)*

Die „*Transcription activator-like effector nucleases*“ (TALENs) gehen auf natürliche Effektorproteine der Bakteriengattung *Xanthomonas* zurück, den sog. „*Transcription activator-like effectors*“ (TALEs). Bei *Xanthomonas* handelt es sich um gram-negative Pflanzenpathogene, die mit Hilfe eines Typ-III-Sekretionssystems ihre Wirte besiedeln und an den Infektionsstellen Chlorose, Zellwucherungen und Nekrose hervorrufen [99; 100]. Hierbei werden über 30 verschiedene Effektormoleküle in die Wirtszelle injiziert, welche diese so umprogrammieren, dass durch ein Herunterfahren des pflanzlichen Immunsystems sowie der Induktion der Zuckersynthese den Bakterien eine starke Vermehrung ermöglicht wird [101]. Um als Aktivator oder Repressor Einfluss auf die Wirtszelle nehmen zu können, müssen die TALEs an die Promotorregion pflanzlicher Gene binden können. Bei der Isolierung und Sequenzierung des ersten dieser Typ-III-Effektorproteine (*avrBs3*) fand man eine zentrale „*repeat*“-Region, bestehend aus Wiederholungen eines hochkonservierten, meist 34 Aminosäuren umfassenden Motivs [102]. Auch in homologen Effektorproteinen konnten diese Motive gefunden werden, jedoch variieren Anzahl, Reihenfolge und Länge dieser Motive und man fand heraus, dass die Aminosäuren an den Positionen 12 und 13 variabel sind („*repeat variable diresidues*“ – RVDs; Abbildung 1.7 B) [103]. Es stellte sich heraus, dass die Gesamtheit dieser tandemartig wiederholten Motive eine völlig neue DNA-Bindedomäne bildet [104; 105]. Durch Sequenzvergleiche konnte man den einzelnen Motiven eine durch die RVDs vermittelte Bindungspräferenz für eine oder mehrere Basen nachweisen [106; 107]. Hiermit war klar, dass diese „*repeat*“-Regionen eine modulare DNA-Bindedomäne darstellen, die theoretisch jede beliebige DNA-Sequenz erkennen und binden kann. Mit Hilfe modularer Klonierungsstrategien konnten TALEs mit gewünschter DNA-Bindespezifität hergestellt und sowohl in pflanzlichen Zellen, als auch in menschlichen Zellen erfolgreich eingesetzt werden [108; 109]. Durch die Fusion der *FokI*-Endonuklease mit TALE-Proteinen entstanden die „*Transcription activator-like effector nucleases*“ (TALENs) und stellten das bis dahin potenteste Werkzeug für das „*genome engineering*“ dar [110]. Da TALENs, genau wie ZFNs, die DNA mit Hilfe der *FokI*-Endonuklease schneiden, müssen auch die TALENs immer als Paar erstellt werden (Abbildung 1.7) [110]. Der große Vorteil der TALENs gegenüber den ZFNs besteht in einer deutlich erhöhten Flexibilität in Bezug auf die Zielsequenz, da hier jede Untereinheit nicht drei, sondern nur eine einzelne Base erkennt. Außerdem lassen sich TALENs dank verschiedener Herstellungsverfahren einfacher und kostengünstiger herstellen [111–114]. Aufgrund dessen sind TALENs bereits bei einer Vielzahl von Organismen, von Viren bis hin zu Säugetieren, erfolgreich zum Einsatz gekommen [86; 87; 115; 116].

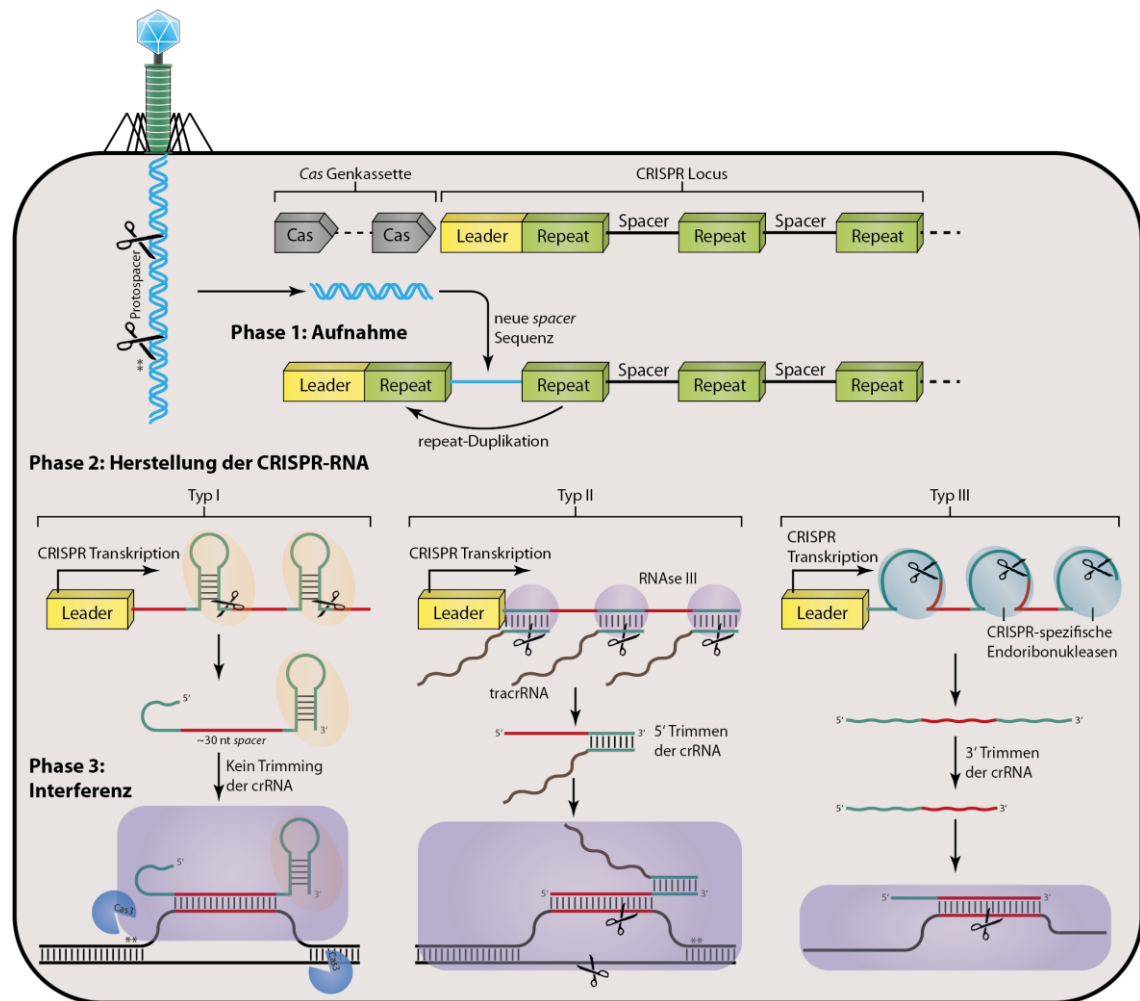


**Abbildung 1.7 | Schematische Darstellung der Funktionsweise verschiedener Werkzeuge für das genome engineering. (A)** Die Zinkfingernukleasen (ZFNs) bestehen aus drei bis vier DNA-Bindedomänen, wobei jede dieser DNA-Bindedomänen drei Nukleotide erkennt und bindet, und aus einer *FokI*-Nuklease Domäne. Um einen Doppelstrangbruch zu erzeugen, müssen zwei ZFNs verwendet werden, da die *FokI* Domänen nur als Heterodimer aktiv sind. **(B)** „Transcription activator like nucleases“ (TALENs) funktionieren ähnlich wie ZFNs, jedoch übernimmt hier eine aus repetitiven Einheiten von 33-35 Aminosäuren bestehende Domäne innerhalb des TALE-Proteins die Bindung an die DNA. Diese erfolgt über die Aminosäuren an den Positionen 12 und 13 („repeat variable diresidue“ (RVD); in rot). **(C)** Das „clustered regularly interspaced short palindromic repeat“ (CRISPR) / Cas9-System setzt sich aus einer ca. 150 Nukleotiden langen „single guide RNA“ (sgRNA) und der Cas9-Nuklease zusammen. Die sgRNA erkennt die 20 bp lange Zielsequenz mittels komplementärer Basenpaarung, während die Cas9-Nuklease das spezifische „protospacer adjacent motif“ (PAM) in der DNA (NGG) erkennt, wodurch es eine Konformationsänderung erfährt und einen Doppelstrangbruch innerhalb der Zielsequenz einführt. (verändert nach: Kim, H; Kim, J: *A guide to genome engineering with programmable nucleases*. Nature Reviews. Genetics 2014, 5:321–334)

### 1.3.3 CRISPR/Cas-System

Nur kurze Zeit nach der Etablierung der TALEN-Technologie wurde mit dem CRISPR/Cas9-System eine effizientere Nachfolgetechnologie etabliert. Obwohl dieses das neueste Werkzeug in dem Bereich des „*genome engineering*“ darstellt, wurde der genomische CRISPR-Locus bereits 1987 zufällig in *Escherichia coli* entdeckt und als durch „*spacer*“ getrennte „*direct repeats*“ mit unbekannter Funktion beschrieben [117]. Spätere Untersuchungen in den 1990er Jahren konnten zeigen, dass homologe Loci auch in vielen anderen Prokaryoten vorkommen [118–123]. Aufgrund der einzigartigen Struktur wurde der Locus als „*Clustered Regularly Interspaced Short Palindromic Repeats*“ (CRISPR) bezeichnet, und man fand heraus, dass dieser Locus von konservierten Genen (CRISPR-assoziierte Gene – Cas-Gene) und einer „*leader*“ Sequenz flankiert wird (Abbildung 1.8) [124]. 2005 stellten verschiedene Forschergruppen fest, dass die „*spacer*“-Sequenzen Homologien zu DNA-Sequenzen verschiedener Viren aufweisen und formulierten die Hypothese, dass Bakterien über ein adaptives Immunsystem verfügen [125–127]. Diese konnte bereits zwei Jahre später experimentell nachgewiesen werden [128]. Die CRISPR/Cas-vermittelte Immunantwort lässt sich in drei Phasen einteilen. Während der initialen Phase erfolgt der Einbau fremden genetischen Materials in den CRISPR-Locus. Ein kurzes Stück der genomischen Fremd-DNA eines Phagen oder Plasmids, der von dem „*protospacer adjacent motif*“ (PAM) flankierte „*protospacer*“ [129], wird unter Beteiligung verschiedener Cas-Proteine herausgeschnitten und integriert zwischen zwei duplizierten „*spacer*“-Elementen in das Genom des Bakteriums (Abbildung 1.8) [128; 130–132]. In der zweiten Phase der Immunantwort erfolgt die Synthese der crRNAs. Hierzu wird ein langes CRISPR-Transkript (pre-crRNA) von dem CRISPR-Locus transkribiert und anschließend durch verschiedene Mechanismen zu reifen crRNAs (CRISPR-RNAs) prozessiert. Insgesamt kann man drei verschiedene CRISPR/Cas-Systeme (vgl. Abbildung 1.8) unterscheiden [133]. Beim Typ-I bildet die repeat-Sequenz durch partielle Rückfaltung eine Haarnadelstruktur aus, welche von den Nukleasen Cas6e und Cas6f erkannt und am 3'-Ende jener Haarnadelstruktur geschnitten wird (siehe Abbildung 1.8) [134–136]. Bei CRISPR/Cas-Systemen des Typs-II hybridisieren „*trans-activating*“ crRNAs (tracrRNAs) an die „*repeat*“-Sequenz der pre-crRNA und bilden so eine RNA-Duplex Struktur aus. Diese wird dann von der zelleigenen RNase III erkannt und geschnitten [137]. In Typ-III-Systemen erfolgt die Prozessierung ähnlich wie in Typ-I Systemen, nur dass hier keine Haarnadelstruktur ausgebildet wird und die Nuklease Cas6 zum Einsatz kommt (siehe Abbildung 1.8) [138]. Während der dritten und letzten Phase bilden die reifen crRNAs mit verschiedenen Cas-Proteinen einen Ribonukleoprotein-Komplex. Dieser ist in der Lage, fremde DNA (oder RNA [139]) über komplementäre Basenpaarungen zwischen „*spacer*“-

Sequenz und der Fremd-DNA zu erkennen, beide Stränge der Fremd-DNA zu schneiden [130] und somit eine Immunität gegen diese Pathogene zu vermitteln [128; 131; 140; 141]. Um eine Reaktion des CRISPR/Cas-Systems gegen das eigene Genom zu schützen, ist dieses so aufgebaut, dass bei einer Komplementarität außerhalb des „*spacers*“ (d.h. sobald auch Teile der benachbarten „*repeat*“-Region an die vermeintliche Fremd-DNA binden) kein Schneiden der DNA mehr möglich ist [142].



**Abbildung 1.8 | Diversität des CRISPR/Cas-vermittelten adaptiven Immunsystems in Prokaryoten.** Der CRISPR-Locus besteht aus tandemartigen „*repeats*“ (grün), welche durch von invasiven genetischen Elementen („*protospacer*“) abstammenden „*spacer*“ getrennt sind und einer „*leader*“-Sequenz (gelb), die als Promotor fungiert. Flankiert wird dieser von den CRISPR-assoziierten Genen (Cas-Gene; grau), die für verschiedene Proteine kodieren, die für die Aufnahme neuer „*spacer*“-Sequenzen (Phase 1), die Herstellung der CRISPR-RNA (Phase 2) und die Ziel-Interferenz (Phase 3) verantwortlich sind. **Phase 1:** Aus dem invasiven genetischen Material werden die durch das „*protospacer adjacent motif*“ (PAM; \*\*) flankierten „*protospacer*“ herausgeschnitten und als neue „*spacer*“-Sequenz zwischen zwei „*repeats*“ in den CRISPR-Locus eingebaut. **Phase 2:** Die langen CRISPR-Transkripte werden durch verschiedene Mechanismen in kurze crRNAs prozessiert. In den Typ-I und -II Systemen kommen CRISPR-spezifische Endonukleasen (orange und blaue Ovale) zum Einsatz, die jeweils acht Nucleotide stromaufwärts der „*spacer*“-Sequenzen schneiden. Im Typ-III System erfolgt ein zusätzliches 3'-Trimming der crRNA. In dem Typ-II System bildet eine „*trans-activating*“ crRNA (tracrRNA) mit der crRNA einen RNA-Duplex, der durch die zelluläre RNaseIII (lila Kreise) erkannt und prozessiert wird. Diese Zwischenstufe wird dann durch ein 5'-Trimming in die reife crRNA umgewandelt. **Phase 3:** In allen Systemen binden die jeweiligen crRNAs an ihre Zielregionen und rekrutieren Cas-Proteine (lila Rechtecke), die Einzel- oder Doppelstrangbrüche in die invasive DNA oder RNA (Typ-III) einführen und eine erneute Infektion verhindern. (verändert nach: Wiedenheft, B; Sternberg, SH; Doudna, JA: *RNA-guided genetic silencing systems in bacteria and archaea*. Nature 2012, 7385:331–338)

Damit war klar, dass das CRISPR/Cas-System eine perfekte Grundlage für ein einfach zu programmierendes und effektives Werkzeug zur zielgerichteten Genommodifikation darstellt. Dieser Aufgabe stellten sich die Mikrobiologin Emmanuelle Charpentier und die US-Strukturbiologin Jennifer Doudna. Sie fusionierten crRNA und tracrRNA zu einer einzelnen chimären „*single-guide*“ RNA (sgRNA) und modifizierten deren „*spacer*“-Sequenz derart, dass diese nun einen Bereich innerhalb der kodierenden Sequenz des Grün Fluoreszierenden Proteins (GFP) erkennt. So konnten sie ein Plasmid, welches das GFP-Gen enthält, erfolgreich *in vitro* schneiden [143] und legten somit den Grundstein für die Verwendung des CRISPR/Cas-Systems für die zielgerichtete Genommodifikation (vgl. Abbildung 1.7). Für ein gezieltes und effektives Ausschalten von Genen bzw. deren Modifikation waren nun nur noch zwei einfach und günstig herzustellende Komponenten, eine sgRNA und das Cas9-Protein bzw. dessen mRNA, notwendig. Aufgrund dessen erschienen bereits fünf Monate nach der Publikation von Charpentier und Doudna zahlreiche weitere Publikationen, die zeigten, dass das System äußerst effektiv in verschiedenen humanen Zelllinien [144], bei Reis- und Weizenpflanzen [145], im Zebrafisch [53; 146], in der Maus [147], in dem Nematoden *Caenorhabditis elegans* [148] und der Taufliege *Drosophila melanogaster* [149] funktioniert. Auch für das sogenannte „*Multiplex genome engineering*“, die gleichzeitige Modifikation mehrerer Gene, ist das CRISPR/Cas-System geeignet [147; 150]. Um ein Gen gezielt ausschalten zu können, ist es nun nur noch notwendig, aus einzelsträngigen Oligonukleotiden eine Matrize für die gewünschte sgRNA herzustellen und diese anschließend zusammen mit einem Cas9-Vektor *in vitro* zu transkribieren und beide RNAs in den zu modifizierenden Organismus einzubringen [151]. Die RNA für die Cas9-Nuklease wird dann *in vivo* durch den Wirtsorganismus translatiert, die sgRNA erkennt und bindet ihre jeweilige Zielsequenz und rekrutiert Cas9, die wiederum das PAM auf der DNA erkennt und dann mit Hilfe der Endonukleasedomänen RuvC und HNH einen Doppelstrangbruch einführt (Abbildung 1.7) [143]. Aufgrund der Tatsache, dass die Zielsequenz durch eine RNA über komplementäre Basenpaarungen erkannt wird, lässt sich so theoretisch jede beliebige Sequenz anvisieren. Jedoch schränkt die Notwendigkeit für das Vorhandensein des „*protospacer adjacent motif*“ (NGG), welches von der Cas9 Nuklease erkannt wird [143], die Auswahlmöglichkeiten für eine geeignete Zielsequenz ein [130; 152]. Zusätzlich schränkt die durch die Verwendung der T7 bzw. SP6 RNA Polymerase für die *in-vitro* Transkription gegebene Anforderung für ein GG- (T7) bzw. GA-Dinukleotid (SP6) am 5'-Ende die Auswahl an Zielsequenzen weiter ein, sodass jede Sequenz der Form 5'-G(G/A)-X<sub>18</sub>-NGG-3' anvisiert werden kann [53]. Trotz der hohen Effizienz und Robustheit hat das CRISPR/Cas-System einen Nachteil. Da dieses System aus Prokaryoten stammt, ist die Zielsequenz mit nur 20 bp für die deutlich größeren eukaryotischen Genome oft nicht spezifisch genug, was zu sogenannten *off-*

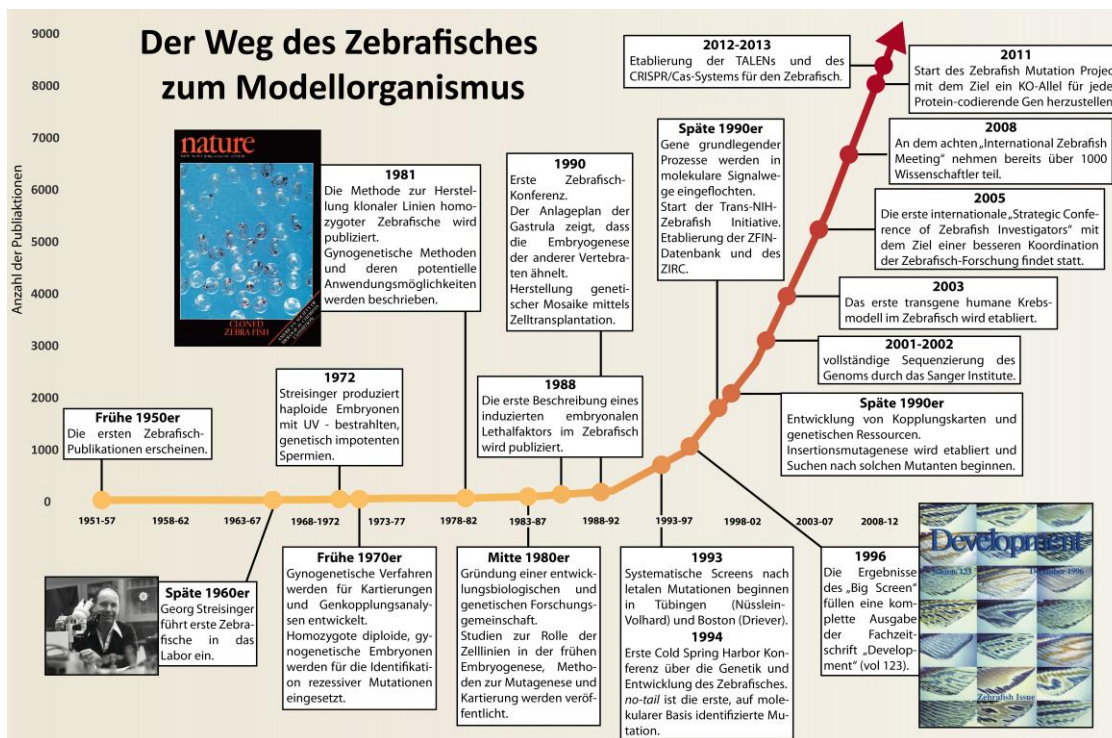
*target* Effekten führen kann [153; 154]. Dies macht eine Analyse möglicher „*off-target*“ Stellen und eine sorgfältige Auswahl der Zielsequenzen sowie deren Genotypisierung nach erfolgter Injektion notwendig. Durch die Verwendung verkürzter sgRNA (tru-sgRNAs) [155] oder gepaarter Cas9-Nickasen, die lediglich einen Einzelstrangbruch in die DNA einführen, können „*off-target*“ Effekte reduziert werden [156; 157].

## 1.4 Der Zebrafisch (*Danio rerio*) als Modellorganismus

Der Zebraäbrbling, *Danio rerio* (früher *Brachydanio rerio*, im Folgenden als Zebrafisch bezeichnet), wurde erstmals bereits 1822 durch Francis Hamilton, einem Chirurgen der British East India Company, beschrieben [158]. Er gehört zu der Familie der karpfenartigen Fische (*Cyprinidae*) [159] und ist in den stehenden Gewässern von Indien und Bangladesch beheimatet [160]. Der Gattungsname *Danio* leitet sich von dem bengalischen Wort „*dhani*“ ab, was so viel bedeutet wie „von dem Reisfeld“ [161].

Obwohl bereits seit Anfang der 1950er Jahre mehrere Publikationen über den Zebrafisch erschienen waren [162], so war es doch Georg Streisinger, der durch seine Arbeiten zur Gynogenese in den späten 1960er und 1970er Jahren (vgl. Abbildung 1.9) den Grundstein für den Zebrafisch als Modellorganismus legte [163–165]. Hinsichtlich der Etablierung von *Danio rerio* als Modellorganismus bestanden jedoch Zweifel, ob Streisingers Ergebnisse in fundamentale Prinzipien übersetzt werden konnten, da zu dieser Zeit noch nicht bekannt war, inwiefern grundlegende Stoffwechselwege und molekulare Mechanismen zwischen stark divergenten Organismen, konserviert sind [164]. Letztendlich führten zwei Faktoren dazu, dass der Zebrafisch seinen Siegeszug als Modellorganismus antreten konnte. Zum einen konnte die Verbindung zwischen der Funktion einzelner Gene und den mutanten Phänotypen in *Drosophila* und *C. elegans* hergestellt werden [164] und zum anderen wurde mehr und mehr klar, dass viele dieser grundlegende Prozesse evolutionär konserviert sind [164; 166; 167].

Der nächste große Meilenstein auf dem Weg des Zebrafisches zum Modellorganismus war der sogenannte „*Big Screen*“ nach embryonalen Mutationen, (C. Nüsslein-Volhard und W. Diever), bei dem ungefähr 2 Millionen durch Mutagenese veränderte Embryonen gesichtet und etwa 4000 mutante Phänotypen gefunden und in einer 481-seitigen Sonderausgabe des Fachmagazins *Development* (Abbildung 1.9) beschrieben wurden [168]. Etwa zur selben Zeit publizierte Charles Kimmel eine äußerst detaillierte Beschreibung der Embryonalentwicklung des Zebrafisches [169].



**Abbildung 1.9 | Meilensteine der Zebrafisch Forschung.** Dargestellt sind die wichtigsten Stationen des Zebrafisches auf dem Weg zum Modellorganismus seit den 1950er Jahren. Auf der vertikalen Achse ist die Anzahl der Publikationen in den jeweils angegebenen Zeiträumen dargestellt. (verändert nach: Grunwald, DJ; Eisen, JS: *Headwaters of the zebrafish -- emergence of a new model vertebrate*. Nat. Rev. Genet. 2002, 9:717–724 & Kinth, P; Mahesh, G; Panwar, Y: *Mapping of zebrafish research: a global outlook*. Zebrafish 2013, 4:510–517)

In den folgenden Jahren wurden Kartierungsexperimente durchgeführt [170–173] und Insertionsmutagenesen [174; 175] sowie „*antisense*“ Methoden [176] angewendet. Im Jahr 2001 wurde das Projekt zur Sequenzierung des Zebrafischgenoms initiiert (Abbildung 1.9) und es zeigte sich, dass für über 70 % der Gene orthologe Gene im Menschen vorhanden sind [177]. Neben der Entwicklungsbiologie spielt der Zebrafisch auch in den Gebieten der Molekularbiologie, Zellbiologie, Genetik, Neurobiologie und der biomedizinischen Forschung eine immer wichtigere Rolle [162; 178; 179]

Zu den Eigenschaften, die den Zebrafisch zu einem der beliebtesten Modellorganismen gemacht haben, gehört neben der geringen Größe, der vergleichsweise kurzen Generationszeit und der hohen Zahl an Nachkommen, auch die optische Transparenz der Eier sowie deren außerhalb des Körpers der Elterntiere stattfindende Befruchtung und Entwicklung [164]. Außerdem haben die breite Verfügbarkeit an genetischen Informationen [177; 180] und Werkzeugen [164; 181; 182; 53] zur Beliebtheit des Zebrafisches beigetragen. Der wohl bedeutendste Vorteil des Zebrafisches ist aber, dass er als Vertebrat dem Menschen phylogenetisch näher steht, als Invertebraten wie *Drosophila* oder *C. elegans*. Gleichzeitig ist es deutlich einfacher, genetische oder embryologische Manipulationen vorzunehmen, als in Säuger-Modellorganismen wie zum Beispiel der Maus [183].



## 1.5 Zielsetzung

Ein Ziel der vorliegenden Arbeit ist die Durchführung von Versuchen zur Etablierung eines „Knock-out“ Stammes für *stk33*, um in Anschluss weitere Informationen über die genaue physiologische Rolle von *stk33* gewinnen zu können. Hierzu wurde der Zebrafisch aufgrund des Status als Modellorganismus und der damit verbundenen Verfügbarkeit von genetischen Informationen und Werkzeugen sowie der einfachen Haltung, der hohen Anzahl an Nachkommen und der ovuliparen Entwicklung und somit einfachen Manipulation der Eier verwendet. Daher soll eine entsprechende Zebrafisch-Anlage etabliert werden, welche die Haltung, Kreuzung und Aufzucht mehrerer verschiedener Populationen ermöglicht. Anschließend soll dann zunächst mit Hilfe der „*Transcription activator-like effector nucleases*“ (TALENs) und später mit Hilfe der im Laufe dieser Arbeit neu verfügbar gewordenen „*Clustered Regularly Interspaced Short Palindromic Repeats*“ (CRISPR) -Technologie *stk33* zunächst in den injizierten Embryonen ausgeschaltet werden. Durch Auskreuzungen der injizierten F<sub>0</sub>-Generation und gezielte Inzucht der daraus hervorgehenden F<sub>1</sub>-Generation soll dann ein vollständig homozygot mutanter „Knock-out“ Stamm (*stk33*<sup>-/-</sup>) etabliert werden.

Da *stk33* bisher, mit Ausnahme der Vögel (*Aves*), in allen Vertebraten nachgewiesen werden konnte, ist ein weiteres Ziel dieser Arbeit die Aufklärung der Fragestellung, ob *stk33* in der Klasse *Aves* tatsächlich fehlt und ob es eventuell auch außerhalb der Wirbeltiere (*Vertebrata*) vorkommt. Für die Analyse der Klasse *Aves* soll zunächst eine geeignete Analysepipeline etabliert werden, welche die Analyse von genomischen sowie transkriptomischen „*next-generation sequencing*“ Datensätzen und bereits assemblierten Genomen bzw. Transkriptomen bezüglich des Vorhandenseins von *stk33* ermöglicht. Die Ergebnisse dieser Analysen sollen dann für ausgewählte Vertreter der Klasse *Aves* mit klassischen PCR-Analysen bestätigt und ggf. ergänzt werden. Dagegen soll die Analyse der vielzelligen Tiere aufgrund der enormen Artenvielfalt und dem damit verbundenen hohen Arbeitsaufwand einer labortechnischen Analyse ausschließlich anhand bereits vorhandener genomischer und transkriptomischer Datensätze erfolgen („*Data Mining*“ oder auch „*Knowledge discovery in databases*“). Die Ergebnisse dieser Untersuchungen sollen Aufschluss über die Verbreitung sowie die Entstehung von *stk33* geben und mögliche Inaktivierungen im Verlauf der Evolution identifizieren.

# 2 | Material & Methoden

## 2.1 Etablierung des Zebrafisches

### 2.1.1 Tierhaltung

Zebrafisch-Embryonen des Stammes AB wurden von dem „*European Zebrafish Resource Center*“ (EZRC) des Karlsruhe Institute of Technology (KIT) bezogen. Die Aufzucht der Tiere erfolgte in 3 l Kunststoff Aquarien, nach ca. vier Wochen wurden diese in mit Kies ausgelegte 80 l Aquarien umgesetzt. Die Wassertemperatur wurde mit Hilfe handelsüblicher Aquarienstabheizern der Fa. Eheim (Deizisau, Deutschland) auf 28 °C eingestellt. Die Umwälzung und Filtration des Wassers erfolgte mehrstufig mittels des Eheim Biopower 160 (Eheim, Deizisau, Deutschland). Die Beleuchtung der Aquarien erfolgte über die Deckenbeleuchtung des Raumes in einem Zyklus aus 14 h Licht und 10 h Dunkelheit.

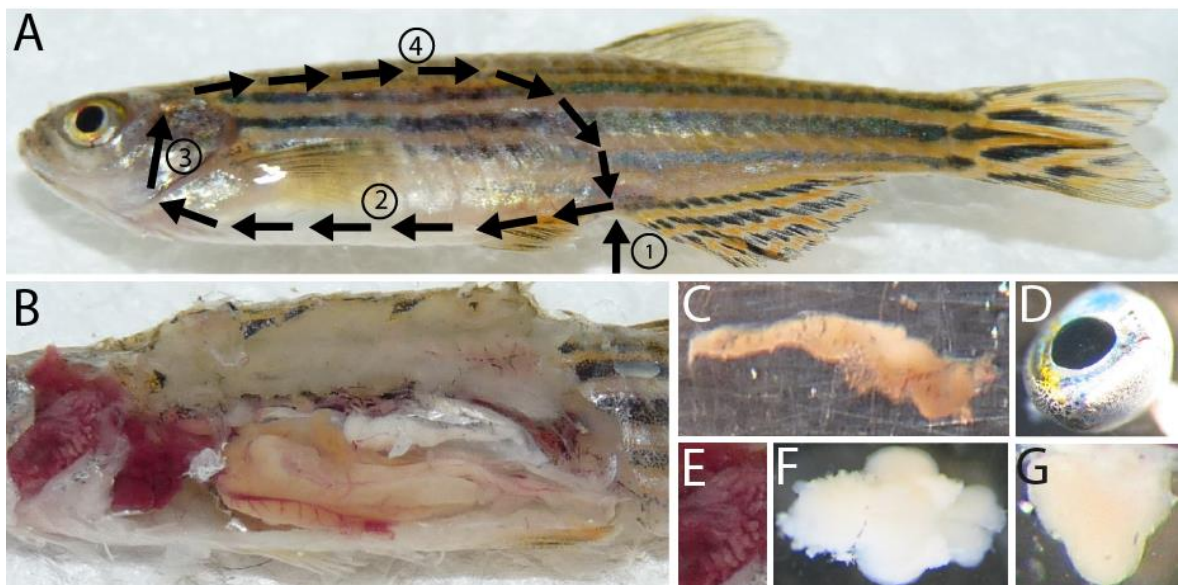
Als Fischfutter wurde hauptsächlich Trockenfutter der Marke TetraMin® (Tetra, Melle, Deutschland) verwendet. Die Fütterung erfolgte zweimal täglich. Wenn Embryonen benötigt wurden, erfolgten zusätzliche Fütterungen mit frisch geschlüpften Nauplien der Gattung *Artemia*. Hierzu wurden am Vortag ca. 4 g Artemia-Eier in 2 l Salzwasser unter ständiger Luftzufuhr angesetzt und bei 28 °C inkubiert. Nach 24 Stunden wurde die Luftzufuhr unterbrochen und die frisch geschlüpften Nauplien konnten vom Boden abgesaugt werden, während die Eierschalen an die Wasseroberfläche trieben. Diese Nauplien wurden dann auf einem Sieb mit VE-H<sub>2</sub>O gereinigt und in eine Petrischale überführt. Eine letzte Trennung der Nauplien von übriggebliebenen Eierschalen erfolgte mit Hilfe einer einseitigen Lichtquelle aufgrund deren positiver Phototaxis. Als Aufzuchtfutter in den ersten ein bis zwei Wochen wurden Protisten der Gattung *Paramecium* verwendet. Diese wurden in 250 ml Zellkulturflaschen, welche mit VE-H<sub>2</sub>O, sechs bis acht aufgekochten Weizenkörnern und Bierhefe versetzt waren, gezüchtet. Nach ca. zwei Wochen erfolgte dann die Umstellung der Jungfische auf *Artemia* Nauplien und Trockenfutter.

### 2.1.2 Gewinnung von Embryonen

Zur Gewinnung von Embryonen wurden schwimmende Ablaichkästen der Fa. JBL (Neuhofen, Deutschland) eingesetzt. Hierbei ist ca. 3 cm über dem Boden ein Gitter angebracht, welches den Laich vor Fraß durch die Elterntiere schützt. Der obere Teil des Ablaichkastens ist durch eine herausnehmbare Trennscheibe in zwei Kammern unterteilt. Am Nachmittag des Vortages wurden dann jeweils zwei Weibchen und ein Männchen getrennt in diese Kammern gesetzt. Am Folgetag wurde ca. 10 min nach Beginn des Lichtzyklus die Trennscheibe herausgenommen und ca. 30 min später konnten die befruchteten Embryonen aus dem Ablaichkasten entnommen und ihrer jeweiligen Verwendung zugeführt werden.

### 2.1.3 Präparation adulter Zebrafische

Die Präparation adulter Zebrafische zur Organentnahme erfolgte nach dem von Gupta und Mullins etablierten Protokoll [184]. Dazu wurde der adulte Zebrafisch in 0,2% Tricain anästhesiert und in -20 °C kalter gesättigter NaCl-Lösung getötet. Zur Freilegung der inneren Organe wurden zunächst Haut und darunterliegendes Muskelgewebe auf der anterioren Seite der Analflosse und anschließend entlang des Bauches bis zum Operculum durchtrennt. Es folgte ein Schnitt durch Haut und Muskelgewebe oberhalb der Kiemen bis zur Analflosse. Nach dem Abheben der Bauchdecke, waren die inneren Organe zugänglich (siehe Abbildung 2.1).



**Abbildung 2.1 | Präparation adulter Zebrafische.** (A) Um die inneren Organe freizulegen, wurde der Zebrafisch zunächst ventral von der Analflosse bis zum Operculum aufgeschnitten. Dieses wurde dann entfernt und ein Schnitt auf der dorsalen Seite bis zu Analflosse legte die (B) inneren Organe frei. So konnten (C) Testes, (D) Auge, (E) Kiemen, (F) Gehirn und (G) Herz präpariert werden.

## 2.2 Molekularbiologische Methoden zur Analyse von DNA

### 2.2.1 Isolierung genomischer DNA

#### 2.2.1.1 Isolierung genomischer DNA mittels Kit

Zur Isolierung genomischer DNA aus Geweben wurde das peqGOLD Tissue DNA Mini Kit der Fa. PeqLab (Erlangen, Deutschland) nach Herstellerangaben verwendet. Die Elution der genomischen DNA erfolgte mit 50 – 100 µl HPLC-H<sub>2</sub>O.

#### 2.2.1.2 Isolierung genomischer DNA mittels HotSHOT-Protokoll

Zwecks Genotypisierung wurde genomische DNA mit Hilfe des HotSHOT-Protokolls isoliert [185; 186]. Hierbei handelt es sich um eine schnelle und kostengünstige Methode zur Isolation von PCR-geeigneter genomischer DNA mit Hilfe von heißem Natriumhydroxid. Hierzu wurde das Ausgangsmaterial in 50 µl einer Natriumhydroxidlösung (25 mM NaOH, 0,2 mM EDTA) bei 98 °C und gelegentlicher Agitation vollständig aufgelöst. Anschließend erfolgte eine Neutralisierung mit 50 µl Tris-HCl (40 mM, pH 5). Um eine Verunreinigung der PCR-Proben mit Zellresten zu verhindern, wurde die so isolierte genomische DNA vor jeder Verwendung bei 3.000 UpM für fünf Minuten zentrifugiert.

#### 2.2.1.3 Isolierung genomischer DNA mittels Proteinase K

Als Alternative zum HotSHOT-Protokoll, wurde genomische DNA mittels eines Proteinase K-haltigen Extraktionspuffers (vgl. 2.6) isoliert [187]. Hierzu wurden zwei bis fünf Embryonen in 50 µl Extraktionspuffer bei 55 °C unter Agitation für zwei bis drei Stunden inkubiert. Anschließend erfolgte eine Inaktivierung der Proteinase K für 10 Minuten bei 99 °C. Für Genotypisierung kompletter, adulter Tiere wurden diese zunächst in flüssigem Stickstoff tiefgefroren und anschließend in einem vorgekühlten Mörser zu einem feinen Pulver verarbeitet. Dieses Pulver wurde dann in 500 µl des Proteinase K-haltigen Extraktionspuffers aufgenommen und ebenfalls für zwei bis drei Stunden bei 55 °C und anschließend 10 Min bei 99 °C inkubiert. Unlösliche Zelltrümmer wurden durch eine fünfminütige Zentrifugation bei 3000 UpM sedimentiert und die Probe bei -20 °C gelagert. Für anschließende PCR-Reaktionen wurden 1-2 µl des Überstandes eingesetzt.

## 2.2.2 Polymerase-Kettenreaktion (PCR)

### 2.2.2.1 Standard PCR

Bei der Polymerase-Kettenreaktion (PCR) handelt es sich um ein von Kary Mullis im Jahre 1986 entwickeltes Verfahren zur *in vitro* Amplifikation eines spezifischen DNA-Abschnittes mit Hilfe einzelsträngiger Oligonukleotide, Desoxyribonukleosidtriphosphaten sowie einer thermostabilen DNA-Polymerase [188]. Als Polymerasen kamen sowohl die GoTaq® DNA Polymerase (Promega, Madison, USA), als auch die Q5® High-Fidelity DNA Polymerase (NEB, Ipswich, USA) zum Einsatz. Alle verwendeten Oligonukleotide wurden von Invitrogen (Carlsbad, USA) und Sigma-Aldrich (St. Louis, USA) synthetisiert. Eine Liste aller verwendeten Primer befindet sich im elektronischen Anhang dieser Arbeit. Die Durchführung der PCRs erfolgte in einem GeneAmp® PCR Systems 9700 Thermocycler (Applied Biosystems, Carlsbad, USA). Beispielhafte PCR-Ansätze für die Q5® bzw. die GoTaq® Polymerase sind in Tabelle 2.1 und Tabelle 2.2 dargestellt. Das entsprechende PCR-Programm kann der Tabelle 2.3 entnommen werden.

Tabelle 2.1 | PCR Ansatz Q5® DNA Polymerase

Komponente	Volumen
5x Q5® Reaktionspuffer	10 µl
5x Q5® High GC Enhancer (optional)	10 µl
dNTPs (10 mM each)	1 µl
Primer, forward (10 µM)	1 µl
Primer, reverse (10 µM)	0,5 µl
Q5® DNA Polymerase (5 U/µl)	1 µl
Template DNA	35,5 µl
HPLC-H <sub>2</sub> O	
<b>Gesamtvolumen</b>	<b>50 µl</b>

Tabelle 2.2 | PCR Ansatz GoTaq® DNA Polymerase

Komponente	Volumen
5x GoTaq® Flexi Puffer	10 µl
MgCl <sub>2</sub> (25 mM)	4 µl
dNTPs (10 mM each)	1 µl
Primer, forward (10 µM)	1 µl
Primer, reverse (10 µM)	1 µl
GoTaq® DNA Polymerase (5 U/µl)	0,2 µl
Template DNA	1 µl
HPLC-H <sub>2</sub> O	31,8 µl
<b>Gesamtvolumen</b>	<b>50 µl</b>

Tabelle 2.3 | PCR-Programm

Schritt	Temperatur	Dauer	Zyklen
<b>Initiale Denaturierung</b>	98 °C	10 min	
<b>Denaturierung</b>	98 °C	30 s	
<b>Annealing</b>	58 - 62 °C	30 s	35 - 42
<b>Elongation</b>	72 °C	60 s/kb	
<b>Finale Elongation</b>	72 °C	10 min	

### 2.2.2.2 „long range“ PCR

Zur Amplifikation von DNA-Abschnitten über 3 kb Länge wurden sog. „long range“ PCR (Ir-PCRs) mit der TaKaRa LA Taq® Hot Start Version (TaKaRa, Otsu, Japan) in einem GeneAmp® PCR Systems 9700 Thermocycler (Applied Biosystems, Carlsbad, USA) durchgeführt. Der Ansatz der Ir-PCRs ist in Tabelle 2.4 gezeigt und das dazugehörige Programm in Tabelle 2.5.

Tabelle 2.4   „long range“ PCR-Ansatz		Tabelle 2.5   „long-range“ PCR-Programm			
Komponente	Volumen	Schritt	Temperatur	Dauer	Zyklen
10x LA PCR Puffer	5 µl	<b>Initiale</b>			
dNTPs (2,5 mM each)	8 µl	<b>Denaturierung</b>	94 °C	10 min	
Primer, forward (10 µM)	2 µl	<b>Denaturierung</b>	94 °C	30 s	
Primer, reverse (10 µM)	2 µl	<b>Annealing</b>	60 - 68 °C	30 s	35 - 40
TaKaRa LA TAq® HS (5 U/µl)	0,2 µl	<b>Elongation</b>	68 °C	60 s/kb	
Template DNA	2 µl	<b>Finale</b>			
HPLC-H <sub>2</sub> O	30,8 µl	<b>Elongation</b>	72 °C	10 min	
<b>Gesamtvolumen</b>	50 µl				

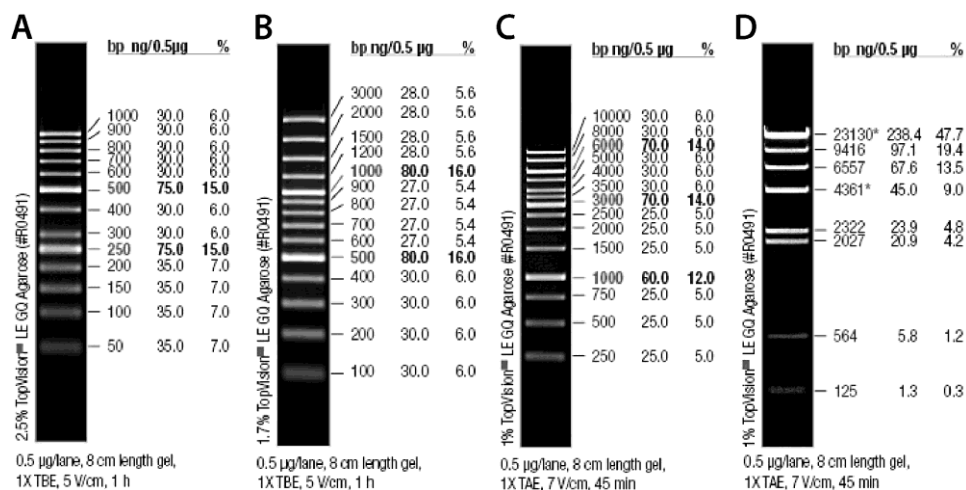
### 2.2.2.3 RACE-PCR

Zur vollständigen Amplifikation von cDNA-Enden wurden sogenannte „rapid amplification of cDNA-ends“ PCR (RACE-PCRs) mit Hilfe des GeneRacer® Kit der Fa. Invitrogen (Carlsbad, USA) nach Herstellerangaben durchgeführt. Anschließend wurden die erhaltenen Amplifikate mittels einer Agarose-Gelelektrophorese (2.2.3) analysiert und mit Hilfe der enthaltenen S.N.A.P.™ Säulen wiedergewonnen und über das TOPO TA Cloning® Kit (Invitrogen, Carlsbad, USA) in den Vektor pCR®4-TOPO® (Vektorkarte siehe Anhang A1) kloniert und sequenziert (siehe 2.2.7).

## 2.2.3 Gelelektrophoretische Auftrennung von DNA

Die Analyse von DNA-Lösungen erfolgte mittels Agarose-Gelelektrophorese in Anlehnung an das Protokoll von Sambrook *et al.* [189]. Hierbei werden Nukleinsäuren umgekehrt proportional zum Logarithmus ihres Molekulargewichtes in einem elektrischen Feld aufgetrennt. Zur Durchführung der Gelelektrophorese wurden sowohl die GENTERPHORESE™-Apparaturen der Fa. Genterprise (Mainz, Deutschland), als auch das „Horizontal System for Submerged Gel Electrophoresis“ der Fa. GIBCO/BRL, heute Life Technologies (Carlsbad, USA), verwendet. Die Agarosekonzentration betrug, je nach Länge der aufzutrennenden Fragmente, 0,8 % - 2,5 % in 1x TBE-Puffer. Die Proben wurden mit 1/6 Volumen eines bromphenolblau-

und xylencyanolhaltigen DNA-Ladepuffers versetzt, um deren Dichte zu erhöhen und zur Markierung der Lauffront. Zur Bestimmung des Molekulargewichtes der in der Probe enthaltenen Fragmente kamen die in Abbildung 2.2 dargestellten Molekulargewichtsstandards zum Einsatz. Die Elektrophorese erfolgte bei 160 V für 20 bis 30 Minuten. Im Anschluss wurde das Gel für zwei Minuten in einer Ethidiumbromidlösung (0,01 mg/ml) gefärbt und auf einem UV-Transilluminator bei einer Wellenlänge von 312 nm mittels des Carestream Gel Logic 112 Imaging System (Carestream Health, Rochester, USA) dokumentiert.



**Abbildung 2.2 | verwendete Molekulargewichtsstandards. (A)** GeneRuler™ 50 bp DNA Ladder, **(B)** GeneRuler™ 100bp plus DNA Ladder, **(C)** GeneRuler™ 1 kb DNA Ladder, **(D)** Lambda/HindIII Marker (alle Thermo Fisher Scientific, Waltham, USA).

## 2.2.4 Wiedergewinnung von DNA-Fragmenten

### 2.2.4.1 Wiedergewinnung durch Elektroelution

Zur Wiedergewinnung von DNA-Fragmenten aus Agarosegelen wurde das gewünschte DNA-Fragment mit Hilfe eines Skalpell herausgeschnitten und in einen ca. 5 cm langen Dialyseschlauch überführt. Nach Zugabe von 300 µl 1 x TBE-Puffer wurde der Dialyseschlauch luftblasenfrei verschlossen und in eine Elektrophoresekammer mit 1 x TBE gelegt. Die Elektroelution erfolgte bei 100 mA für ca. 15 min. Der Fortschritt der Elution wurde mit Hilfe einer UV-Handlampe kontrolliert. Wenn die DNA vollständig aus dem Agarosegel gewandert war, wurde die Polung für eine Minute umgekehrt, um am Dialyseschlauch haftende DNA wieder in Lösung zu bringen. Die gelöste DNA wurde anschließend in ein 1,5 ml Reaktionsgefäß überführt und der Dialyseschlauch nochmals mit 200 µl 1 x TBE-Puffer gespült. Diese Lösung wurde dann mit dem Eluat vereinigt und zwecks Aufreinigung und Konzentration der eluierten DNA wie in 2.2.5.1 beschrieben gefällt.

### **2.2.4.2 Wiedergewinnung mittels Kit**

Alternativ zur Wiedergewinnung durch Elektroelution wurde das GenElute™ Gel Extraction Kit der Fa. Sigma-Aldrich (St. Louis, USA) nach Herstellerangaben eingesetzt. Die Elution der DNA erfolgte in 30 µl 70 °C warmen HPLC-H<sub>2</sub>O.

## **2.2.5 Aufreinigung von DNA**

### **2.2.5.1 Fällung von DNA**

Zur Fällung wurde die DNA-Lösung mit 1/10 Volumen 10 x Dialysepuffer und 2,5 Volumen Ethanol abs. versetzt und invertiert. Es folgte wahlweise eine Inkubation bei -25 °C für mindestens 1 Stunde oder bei -80 °C für 30 min. Die Präzipitation der gefällten DNA erfolgte durch Zentrifugation in der Kühlzentrifuge 5804R der Fa. Eppendorf (Hamburg, Deutschland) bei 14.000 UpM und 4 °C für 30 min. Anschließend wurde der Überstand verworfen und das Pellet mit 500 µl 70 %-igem Ethanol gewaschen und erneut für 10 min bei 14.000 UpM und 4 °C zentrifugiert. Der Überstand wurde erneut verworfen und das Pellet in der Vakuumzentrifuge Concentrator 5301 der Fa. Eppendorf (Hamburg, Deutschland) getrocknet und anschließend in HPLC-H<sub>2</sub>O gelöst.

### **2.2.5.2 Aufreinigung von PCR-Produkten**

Um PCR-Produkte direkt sequenzieren zu können, mussten zunächst verbliebene Primer und dNTPs verdaut werden. Dies erfolgte durch eine Inkubation mit 5 U Exonuklease I und 0,9 U Shrimp Alkalischer Phosphatase (beide Fermentas, St.-Leon-Rot, Deutschland) für 25 min bei 37 °C. Zur Inaktivierung der Enzyme erfolgte eine Inkubation bei 72 °C für 15 min.

## **2.2.6 Klonierung von PCR-Produkten**

### **2.2.6.1 Ligation**

Für die Ligation wurde die T4 DNA Ligase (Fermentas, St.-Leon-Rot, Deutschland) verwendet. Als Vektoren kamen sowohl das pGEM® T Easy Vector System (Promega, Madison, USA), als auch der Vektor pUC19 (Fermentas, St.-Leon-Rot, Deutschland) zum Einsatz (Vektorkarten siehe Anhang; Abbildung A 1 und Abbildung A 2). Ein typischer Ligationsansatz ist in Tabelle 2.6 dargestellt. Die Ligation erfolgte bei 4 °C über Nacht. Anschließend wurde der Ligationsansatz gefällt (siehe 2.2.5.1) und in 20 µl HPLC-H<sub>2</sub>O gelöst.



Tabelle 2.6 | Ligationsansatz

Komponente	Volumen
DNA-Lösung	3 $\mu$ l
Vektor (5 ng/ $\mu$ l)	1 $\mu$ l
10 x DNA Ligase Puffer	1 $\mu$ l
T4 DNA Ligase (3 U/ $\mu$ l)	1 $\mu$ l
HPLC-H <sub>2</sub> O	4 $\mu$ l
Gesamtvolumen	10 $\mu$ l

### 2.2.6.2 Transformation

Die Transformation der Ligationen in kompetente *Escherichia coli* Zellen des Stammes DH10B erfolgte per Elektroporation mit Hilfe des MicroPulser™ (BioRad, Hercules, USA). Hierfür wurden 5  $\mu$ l des gefällten Ligationsansatzes zu den eiskalten *E. coli* Zellen pipettiert, für eine Minute auf Eis inkubiert und anschließend bei 1,8 kV transformiert. Die transformierten Zellen wurden sofort in 1 ml auf 37 °C vorgewärmtem L-Medium aufgenommen und für 30 Minuten bei 37 °C inkubiert. Anschließend wurden 200  $\mu$ l des Transformationsansatzes auf einer Agarplatte ausplattiert und über Nacht bei 37 °C inkubiert.

### 2.2.6.3 Plasmidisolierung

Nach erfolgter Transformation wurden rekombinante Klone identifiziert und in 5 ml L-Medium, welches mit 50  $\mu$ l einer Ampicillin Stammlösung (10 mg/ml) versetzt wurde, aufgenommen. Die Anzucht der Bakterien erfolgte über Nacht bei 37 °C unter ständiger Agitation. Zur Isolierung der rekombinanten Plasmide wurden das GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich, St. Louis, USA) und das peqGOLD Plasmid Miniprep Kit der Fa. PeqLab (Erlangen, Deutschland) nach Herstellerangaben verwendet. Die Elution der Plasmide erfolgte in 50  $\mu$ l 70 °C warmen HPLC-H<sub>2</sub>O. Zur Bestimmung der Integratgröße wurden pGem® T Easy-Plasmide mit EcoRI (Fermentas, St.-Leon-Rot, Deutschland) und pUC19-Plasmide mit EcoRI und HindIII (beide Fermentas, St.-Leon-Rot, Deutschland) verdaut und elektrophoretisch aufgetrennt (2.2.2.3).

### 2.2.7 Sanger-Sequenzierung

Für die Sanger-Sequenzierung [190] wurden 50-100 ng der aufgereinigten PCR-Produkte (siehe 2.2.5.2) bzw. 400-700 ng der Plasmide mit je 1  $\mu$ l des jeweiligen Primer (10  $\mu$ M) bei der Fa. StarSEQ (Mainz, Deutschland) sequenziert. Die Bewertung der Qualität der erhaltenen Sequenzierungen erfolgte mit Hilfe des Programmes FinchTV v1.4.0 (Geospiza, Seattle, USA).

### **2.2.8 Konzentrations- und Reinheitsbestimmung von DNA**

Um eine genauere Aussage über Reinheit und Konzentration von DNA-Lösungen machen zu können, wurde diese photometrisch aus einem Probenvolumen von 1 µl mit dem Nanodrop® ND-1000 (PeqLab, Erlangen, Deutschland) gemessen. Die Reinheit kann mit Hilfe des Quotienten der Absorption bei 260 nm für Nukleinsäuren und 280 nm für Proteine ( $A_{260/280}$ ) bestimmt werden. Für reine DNA liegt dieser bei 1,8.

## **2.3 Molekularbiologische Methoden zur Analyse von RNA**

### **2.3.1 Isolierung von Total-RNA**

Zur Isolation von Total-RNA wurde das RNeasy Mini Kit (Qiagen, Hilden, Deutschland), das High Pure RNA Tissue Kit (Roche, Penzberg, Deutschland) und das PureLink® RNA Mini Kit (Life Technologies, Carlsbad, USA) nach Herstellerangaben verwendet. Das Ausgangsmaterial wurde in dem jeweiligen Lysepuffer aufgenommen und mit Hilfe einer Kanüle (Außendurchmesser 0,7 mm) homogenisiert. Sofern ein DNase-Verdau nicht im Protokoll des Herstellers vorgesehen war, erfolgte dieser zusätzlich auf der Silikamembran (DNaseI, NEB, Ipswich, USA). Die Elution der Total-RNA erfolgte mit 30 µl 70 °C warmen HPLC-H<sub>2</sub>O. Die isolierte Total-RNA wurde bei -80 °C gelagert.

### **2.3.2 Konzentrations- und Reinheitsbestimmung von RNA**

#### **2.3.2.1 Qubit**

Die fluorimetrische Konzentrationsbestimmung von RNA-Proben mit dem Qubit® (Invitrogen, Carlsbad, USA) erfolgte unter Verwendung des Qubit® RNA HS Assay Kit nach Herstellerangaben. Es wurde 1 µl Probenvolumen zur Messung eingesetzt.

#### **2.3.2.2 Bioanalyser**

Um sowohl eine quantitative, als auch eine qualitative Aussage [191] über die RNA treffen zu können, wurden Messungen mit dem Agilent 2100 Bioanalyser (Agilent, Santa Clara, USA) durchgeführt. Diese erfolgten mit Hilfe des RNA 6000 Nano Kits nach Herstellerangaben.

### 2.3.2.3 RNA-Gelelektrophorese

Die gelelektrophoretische Auftrennung von RNA erfolgte in einem denaturierenden Agarosegel. Hierzu wurden 0,3 g Agarose in 18,75 ml DEPC-H<sub>2</sub>O gelöst und aufgeköcht. Anschließend wurden 2,5 ml 10 x MOPS-Puffer und 3,75 ml 37 % Formaldehyd zugegeben. Die RNA-Proben wurden mit HPLC-H<sub>2</sub>O auf ein Volumen von mindestens 3 µl aufgefüllt und mit 1 Vol. des ethidiumbromid-haltigen 2 x RNA Ladepuffer (Thermo Fisher Scientific, Waltham, USA) versetzt. Vor dem Auftragen wurden die Proben für 10 min bei 70 °C denaturiert. Die Elektrophorese erfolgte bei 160 V für 20 bis 30 Minuten. Als Laufpuffer diente 1 x MOPS-Puffer. Die Dokumentation erfolgte auf einem UV-Transilluminator bei einer Wellenlänge von 320 nm mit Hilfe des Carestream Gel Logic 112 Imaging System (Carestream Health, Rochester, USA).

### 2.3.3 Wiedergewinnung von RNA-Fragmenten mittels Elektroelution

Die Wiedergewinnung von RNA-Fragmenten aus Agarosegelen erfolgte wie in 2.2.4.1 beschrieben. Der Dialyseschlauch wurde jedoch vor Versuchsbeginn für 10 Minuten mit RNase Away® behandelt und als Laufpuffer kam 1 x MOPS-Puffer zum Einsatz.

### 2.3.4 cDNA-Synthese

Die cDNA wurde mit Hilfe des SuperScript® III Kits der Fa. Invitrogen (Carlsbad, USA) nach Herstellerangaben synthetisiert. Die Grundlage für die reverse Transkription bildete 1 µg Total-RNA.

### 2.3.5 *in vitro* Transkription

Zur Herstellung der TALEN mRNAs sowie der Cas9 mRNA wurde das T7 mScript® Standard mRNA Production System der Fa. Cellscript (Madison, USA) nach Herstellerangaben verwendet. Die Aufreinigung der mRNAs erfolgte mittels Ammoniumacetatfällung nach Angaben des Herstellers.

Die *in vitro* Transkription der sgRNAs erfolgte mit Hilfe des T7 RNA Polymerase-Plus™ Enzyme Mix (Life Technologies, Carlsbad, USA) nach Herstellerangaben. Ein exemplarischer Ansatz ist in Tabelle 2.7 gezeigt. Die Transkription erfolgte bei 37 °C für ca. 3 Stunden.

Tabelle 2.7 | Ansatz für die *in vitro* Transkription der sgRNA

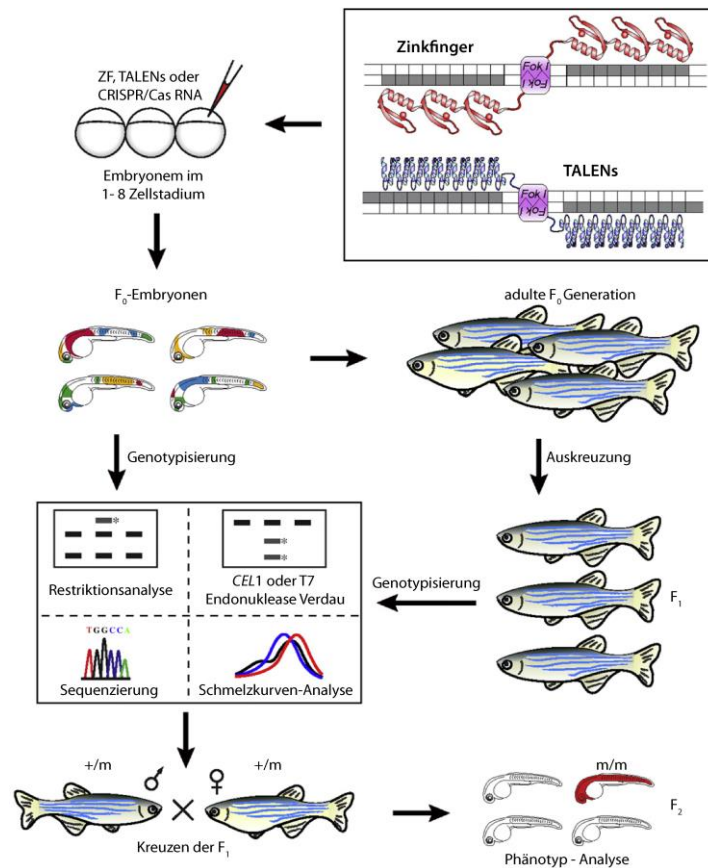
Komponente	Volumen
10x Transkriptions-Puffer	2 µl
rNTPs (10mM)	4 x 3 µl
Template DNA (1 µg/µl)	1 µl
T7 Enzym Mix (20 U/µl)	1 µl
HPLC-H <sub>2</sub> O	4 µl
Gesamtvolumen	20 µl

### 2.3.6 Fällung von RNA

Zur Aufreinigung und Aufkonzentrierung von RNA-Proben wurden Ethanolfällungen durchgeführt. Hierzu wurden diese mit HPLC-H<sub>2</sub>O auf ein Volumen von mindestens 100 µl gebracht, mit  $\frac{1}{10}$  Vol. 3 M Natriumacetat und 2,5 Vol. eiskaltem Ethanol abs. versetzt und invertiert. Die Fällung erfolgte für ca. 2 Stunden bei -25 °C. Darauf folgte eine Zentrifugation für 30 min bei 14.000 UpM und 4 °C um die gefällte RNA zu pelletieren. Dieses Pellet wurde dann mit 500 µl 70 % Ethanol gewaschen und erneut für 10 min bei 14.000 UpM und 4 °C zentrifugiert. Anschließend wurde das RNA-Pellet getrocknet und in HPLC-H<sub>2</sub>O gelöst.

## 2.4 genome engineering

Der Begriff „*genome engineering*“ fasst alle Strategien und Techniken zur gezielten Mutagenese von Genomen zusammen. Zu diesen Techniken gehören u.a. die Zinkfinger nukleasen, die TALENs und das CRISPR/Cas-System. Zwar unterscheiden sich diese Systeme sehr stark in der Art Ihrer Herstellung (vgl. hierzu 2.4.1 und 2.4.2), jedoch ist der darauf folgende Arbeitsablauf zur Erzeugung eines gentechnisch veränderten Organismus identisch (siehe Abbildung 2.3). Hierzu wird zunächst das jeweilige RNA-Gemisch in Embryonen injiziert. Ein Teil dieser injizierten Embryonen wird dann wie unter 2.4.4 beschrieben genotypisiert. Wenn diese Genotypisierung eine erfolgreiche Mutation an der Zielsequenz anzeigt, werden diese F<sub>0</sub>-Embryonen aufgezogen und anschließend ausgekreuzt. Die daraus resultierende F<sub>1</sub>-Generation wird dann noch im Embryonalstadium auf Ihre Heterozygotie überprüft (vgl. 2.4.4) und ggfs. aufgezogen. Die heterozygoten Fische dieser F<sub>1</sub> Generation werden dann untereinander gekreuzt, sodass in der daraus resultierenden F<sub>2</sub>-Generation homozygot mutante Embryonen erzeugt werden können (siehe Abbildung 2.3).



**Abbildung 2.3 | gezielte Mutagenese mittels ZFNs, TALENs und des CRISPR/Cas-Systems.** Die jeweiligen RNA-Gemische werden in Embryonen injiziert und diese anschließend genotypisiert. Erfolgreich injizierte Embryonen werden aufgezogen und ausgekreuzt. Die daraus entstehende F<sub>1</sub> Generation wird dann noch im Embryonalstadium auf Ihre Heterozygotie überprüft und aufgezogen. Durch das Kreuzen dieser F<sub>1</sub> Generation entstehen dann 25 % mutante Embryonen in der F<sub>2</sub> Generation. (verändert nach Huang, P; Zhu, Z; Lin, S; Zhang, B: *Reverse Genetic Approaches in Zebrafish*. Journal of Genetics and Genomics 2012, 9:421–433 [192])

### 2.4.1 TALENs

Die Auswahl geeigneter Zielsequenzen für die TALENs erfolgte mit Hilfe des webbasierten „TAL Effector Nucleotide Targeter 2.0“ der Cornell University [193; 111]. Die Konstruktion der TALENs erfolgte nach dem von Sanjana *et al.* publizierten Protokoll [194] und die dazu notwendigen Plasmide wurden von dem Plasmid Repository addgene (Cambridge, USA) bezogen. Dieses Protokoll basiert auf einer Assemblierung der einzelnen Monomere durch eine hierarchische Ligationsstrategie. Hierbei werden zunächst die einzelnen Monomere durch PCR-Amplifikation mit Hilfe spezifischer Primer mit Typ-II Restriktionsschnittstellen versehen, welche die spätere Position des Monomers im TALEN-Vektor festlegen. Im nächsten Schritt werden, je nach Zielsequenz, die entsprechenden Monomere in einer „Golden Gate“ Reaktion, einer Kombination aus Restriktionsverdau und gleichzeitiger Ligation, zu je drei zirkulären Hexameren zusammengesetzt. Diese wurden anschließend erneut per PCR amplifiziert und in einer zweiten „Golden Gate“ Reaktion direkt in den entsprechenden TALEN-Vektor kloniert.

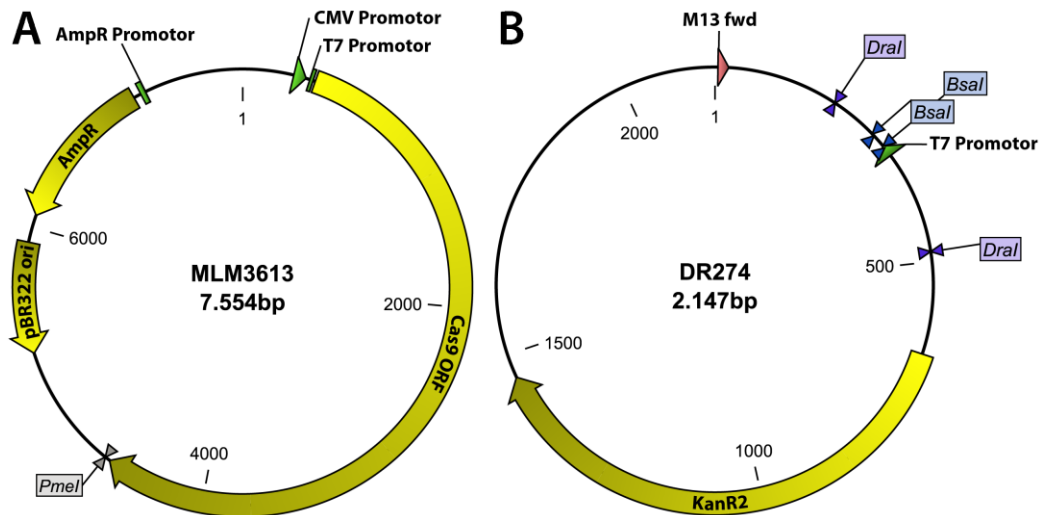
Zwecks Überprüfung der hergestellten TALEN-Vektoren wurden diese transformiert (vgl. 2.2.6.2) und anschließend über einen Restriktionsverdau mit *Sall* charakterisiert, da die Kolonie-PCR wie sie von Sanjana *et al.* [194] verwendet wurde, nur unzuverlässig funktionierte. Hierbei entstehen bei einem korrekt assemblierten TALEN-Vektor drei Restriktionsfragmente mit einer Größe von 2,3 kb, 2,7 kb und 3,3 kb. Positiv getestete Vektoren wurden anschließend zur genaueren Überprüfung sequenziert (siehe 2.2.7).

Die *in vitro* Transkription der korrekt assemblierten TALEN Vektoren erfolgte wie unter 2.3.5 beschrieben. Die hergestellten TALEN-mRNAs wurden dann sowohl auf einem denaturierenden RNA Gel (2.3.2.3), als auch auf dem Agilent 2100 Bioanalyzer (2.3.2.2) analysiert. Die Mikroinjektion erfolgte wie unter 2.4.3 beschrieben mit einer Injektionslösung, die variierende TALEN-mRNA Konzentrationen von 5 ng/μl bis 100 ng/μl enthielt.

#### 2.4.2 CRISPR/Cas

Die Auswahl geeigneter Zielsequenzen für das CRISPR/Cas-System erfolgte durch das webbasierte Programm „ZiFIT Targeter Version 4.2“ des Zinc Finger Consortium [195; 196]. Hierbei wurden auch potenzielle „*off target sites*“ identifiziert. Die Konstruktion der Cas9 mRNA und der sgRNA erfolgte nach dem von Hwang *et al.* publizierten Protokoll [53]. Hierbei wurden die in Abbildung 2.4 dargestellten Vektoren verwendet. Der Vektor MLM3613 kodiert hierbei für die Cas9 Nuklease. Dieser wurde mit Hilfe des Restriktionsenzym *PmeI* linearisiert, gefällt und für die *in vitro* Transkription (vgl. 2.3.5) eingesetzt. Für die *in vitro* Transkription der sgRNA wurden zunächst zwei kompatible Oligonukleotide für die jeweilige Zielsequenz mit Hilfe des „ZiFIT Targeter Version 4.2“ erstellt und von der Fa. Sigma Aldrich (St. Louis, USA) synthetisiert. Diese wurden dann in Anwesenheit von 1 x Annealing Puffer aneinandergelagert (siehe Tabelle 2.8) und in den *BsaI* verdauten Vektor DR274 kloniert. Dieser wurde dann transformiert (2.2.6) und zur Überprüfung sequenziert (2.2.7). Anschließend erfolgte ein Restriktionsverdau mit *DraI*, sodass das Template für die sgRNA aus dem Vektor herausgeschnitten wurde (siehe Abbildung 2.4). Der Restriktionsverdau wurde dann über eine Agarose-Gelelektrophorese (2.2.3) aufgetrennt und das DNA-Fragment, welches das Template für die sgRNA enthält, per Elektroelution wiedergewonnen (2.2.4.1) und für die *in vitro* Transkription eingesetzt (2.3.5). Alternativ zu diesem Verfahren für die Herstellung der sgRNA wurden Oligonukleotide, die für die vollständige Sequenz der sgRNA kodieren, erstellt und synthetisiert (Sigma-Aldrich, St. Louis, USA). Diese Oligonukleotide wurden dann ebenfalls unter Anwesenheit von 1 x Annealing Puffer aneinandergelagert (siehe Tabelle 2.8), gefällt und direkt für die *in vitro* Transkription eingesetzt (2.3.5). Die Analyse der hergestellten RNAs

erfolgte über eine denaturierende Agarose-Gelelektrophorese (2.3.2.3) und die Konzentration der RNA-Lösungen wurde mit Hilfe des Qubit® (Invitogen, Carlsbad, USA) bestimmt (2.3.2.1). Die Injektionslösung wurde mit 150 ng/μl der Cas9 mRNA und 6,25 ng/μl der sgRNA angesetzt und mit 0,05 % Phenolrot versetzt.



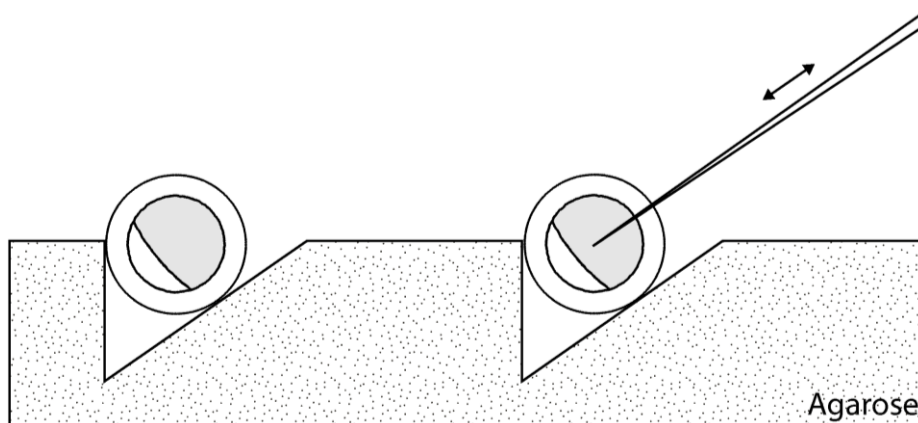
**Abbildung 2.4 | Übersicht über die verwendeten CRISPR/Cas-Vektoren.** Dargestellt sind (A) der Vektor MLM3613, der für die Cas9 Nuklease kodiert und (B) der Vektor DR274, der für die sgRNA kodiert. Funktionale genetische Elemente sind als gelbe Pfeile, Promotoren als grüne Pfeile dargestellt. Die für die Herstellung der CRISPR/Cas-RNAs notwendigen Restriktionschnittstellen sind ebenfalls dargestellt.

### 2.4.3 Mikroinjektion

Das Einbringen der TALEN-mRNAs und der CRISPR/Cas-RNAs in Zebrafisch-Embryonen erfolgte per Mikroinjektion. Hierzu wurden die Mikroinjektionsnadeln aus dickwandigen Glaskapillaren (1,00 x 0,58 x 80 mm; GB100F-8P) der Fa. Science Products (Hofheim, Deutschland) mit Hilfe des P-97 Micropipette Pullers (Sutter Instruments, Novato, USA) hergestellt (Heat: 645; Pull: 60; Velocity: 80). Am Vorabend der Injektion wurden adulte Fische wie in 2.1.2 beschrieben angesetzt.

Unmittelbar vor der Injektion wurde die jeweilige Injektionslösung frisch angesetzt und die Injektionsnadel mit einem Skalpell aufgeschnitten, sodass eine passende Öffnung entstand. Hierbei war es von essentieller Bedeutung, dass der Schnitt etwa 1 bis 2 mm in einem Winkel von ca. 45° oberhalb der Nadelspitze erfolgte. Wurde der Schnitt zu tief angesetzt, war die Nadel zu fein um das Chorion des Embryos zu durchdringen. Wurde der Schnitt zu hoch angesetzt, war die Nadel zu grob, sodass der Embryo durch die Injektion zu stark verletzt wurde und das injizierte Volumen aufgrund der großen Öffnung nicht mehr regulierbar war.

Zur Injektion wurden die Embryonen auf einer Agaroseplatte mit keilförmigen Vertiefungen aufgereiht (siehe Abbildung 2.5). Die Führung der Nadel erfolgte mit Hilfe eines Mikromanipulators. Die Injektion erfolgte dann mittels des FemtoJet® express der Fa. Eppendorf (Hamburg, Deutschland). Der Injektionsdruck ( $P_i$ ) wurde so gewählt, dass das Injektionsvolumen ca. 10 % des Volumens des Embryos entsprach. Dieser lag hierbei zwischen 0,1 und 1,0 psi bei einer Injektionsdauer von 0,1 s. Der Kompensationsdruck ( $P_c$ ), welcher ein Einströmen von Medium und Dottermasse in die Nadel verhinderte, lag zwischen 1,0 und 4,0 psi. Die Injektionslösung wurde durch das Chorion direkt in den Dottersack der Embryonen, die sich im 1- bis 4-Zell-Stadium befanden, injiziert (siehe Abbildung 2.5). Nach erfolgter Injektion wurden die injizierten Embryonen in 1x E3-Medium aufgenommen und für 48 Stunden bei 28,5 °C inkubiert und anschließend genotypisiert (vgl. 2.4.4). Als Kontrolle wurden 20 bis 30 uninjizierte Embryonen mitgeführt.



**Abbildung 2.5 | Schematische Darstellung der Mikroinjektion.** Die Embryonen wurden auf einer Agaroseplatte aufgereiht und durch das Chorion in den Dottersack injiziert.

#### 2.4.4 Genotypisierung

Die Genotypisierung der injizierten Embryonen wurde mit Hilfe der T7 Endonuklease I durchgeführt. Dazu wurde zunächst genomische DNA aus 48 hpf Embryonen wie in 2.2.1.2 und 2.2.1.3 beschrieben isoliert und anschließend die zu genotypisierende Region per PCR amplifiziert (vgl. 2.2.2.1). 10 µl dieses PCR-Produktes wurden dann mit 1,2 µl 10x NEBuffer 2 versetzt und nach dem in Tabelle 2.8 dargestellten Programm denaturiert und anschließend renaturiert. Hierbei bilden sich Heteroduplexe aus, sofern ein mutanter und ein wildtypischer Strang hybridisieren. Im Anschluss an die Renaturierung wurde 1 µl der T7 Endonuklease I (NEB, Ipswich, USA) zugegeben und es erfolgte eine Inkubation bei 37 °C für 15 Minuten. Hierbei erkennt die T7 Endonuklease I diese Heteroduplexe und führt an diesen Stellen



Doppelstrangbrüche ein. Der Reaktionsansatz wurde dann wie unter 2.2.5.1 beschrieben gefällt und elektrophoretisch aufgetrennt (siehe 2.2.3).

Die Genotypisierung der aus den Auskreuzungen entstandenen F<sub>1</sub> Embryonen erfolgte sowohl mit Hilfe der T7 Endonuklease I, als auch durch direkte Sanger-Sequenzierung (2.2.7) der PCR-Produkte. Adulte Tiere wurden zwecks Genotypisierung unter flüssigem Stickstoff mit Hilfe von Mörser und Pistill zu einem feinen Pulver verarbeitet, welches dann als Ausgangsmaterial für die Isolierung genomischer DNA wie unter 2.2.1.3 beschrieben diente. Im Anschluss erfolgte dann zunächst die Amplifizierung der Zielregion und die eigentliche Genotypisierung erfolgte dann, wie bereits beschrieben, mittels des T7E1-Assay.

Die genauere Charakterisierung der genomischen Zielregion erfolgte durch Amplifikation (2.2.2.1) und Klonierung (2.2.6) der Zielregion und der anschließenden Sequenzierung dieser Klone (2.2.7).

**Tabelle 2.8 | Programm zur De- und Renaturierung.**

Temperatur	Zeit	$\Delta T$
98 °C	10 min	
98 °C - 85 °C		-2,0 °C/s
85 °C	1min	
85 °C - 75 °C		-0,3 °C/s
75 °C	1min	
75 °C - 65 °C		-0,3 °C/s
65 °C	1min	
65 °C - 55 °C		-0,3 °C/s
55 °C	1min	
55 °C - 45 °C		-0,3 °C/s
45 °C	1min	
45 °C - 35 °C		-0,3 °C/s
35 °C	1min	
35 °C - 25 °C		-0,3 °C/s
25 °C	1min	
8 °C	$\infty$	

## 2.5 Bioinformatische Methoden

### 2.5.1 Konvertierung der SRA-Datensätze

Zur Analyse der Expression und der genomischen Präsenz von *stk33* in der Klasse *Aves* wurden sowohl transkriptomische als auch genomische Next-Generation Sequencing Datensätze aus dem NCBI „*Sequence Read Archive*“ (SRA) heruntergeladen (<http://www.ncbi.nlm.nih.gov/sra>). Eine Übersicht über alle in dieser Arbeit verwendeten Next-Generation Sequencing Datensätze ist in Tabelle 2.9 dargestellt. Die Datensätze befanden sich zu diesem Zeitpunkt im sra-Dateiformat. In diesem Dateiformat sind die forward und reverse Sequenzierungen zu einer einzelnen Sequenz zusammengefügt. Daher mussten diese zunächst getrennt und anschließend in das fastq-Dateiformat umgewandelt werden, um einen Import in die CLC Genomics Workbench v6.5.1 (CLC bio, Aarhus, Dänemark) zu ermöglichen. Außerdem wurden die fastq-Dateien mit gzip gepackt. All diese Schritte wurden mit Hilfe des sra toolkit v2.3.5-2 [197] durchgeführt. Die so erhaltenen fastq.gz Dateien wurden dann zur weiteren Bearbeitung (siehe 2.5.2) in die CLC Genomics Workbench (CLC bio, Aarhus, Dänemark) importiert.

### 2.5.2 Filterung der Datensätze

Nach der Umwandlung der sra-Dateien in das fastq.gz-Format wurden diese Datensätze über die Funktion „*NGS Import*“ unter Berücksichtigung der jeweiligen minimalen und maximalen Abstände der Sequenzen in die CLC Genomics Workbench (CLC bio, Aarhus, Dänemark) importiert. Zur Entfernung von Sequenzen mit schlechter Qualität und eventuell noch vorhandener Adapter-Sequenzen wurde die Funktion „*Trim Sequences*“ gewählt. Hierbei wurde ein Phred-Wert von 20, was einer Fehlerwahrscheinlichkeit von 0,01 entspricht, als Grenzwert verwendet. Außerdem durften die Sequenzen nicht mehr als eine nicht eindeutige Base (N) enthalten. Alle Sequenzen mit einer Länge von weniger als 15 Basen wurden verworfen.

**Tabelle 2.9 | Übersicht über alle für die Analyse der Klasse Aves verwendeten Next-Generation Sequencing Datensätze.** Die SRA Studien SRX736627, SRX738984 und SRX738987 wurden am IMSB Mainz durchgeführt. Alle anderen Datensätze wurden aus dem NCBI Sequence Read Archive heruntergeladen.

Organismus	Herkunft d. Sequenzen	Gewebe	SRA Studie	Plattform	Anzahl d. Sequenzen	Layout
Afrikanischer Vogelstrauß ( <i>Struthio camelus</i> )	transkriptomisch	Cerebellum	SRX736627	Illumina HiSeq 2000	70,355,488	100 bp paired
Afrikanischer Vogelstrauß ( <i>Struthio camelus</i> )	transkriptomisch	Testis	SRX738984	Illumina HiSeq 2000	32,268,702	100 bp paired
Afrikanischer Vogelstrauß ( <i>Struthio camelus</i> )	genomisch	-	SRX334065 SRX334066 SRX334067 SRX334068 SRX334072	Illumina HiSeq 2000	210,096,016 237,908,622 183,703,328 170,644,024 160,894,606	100 bp paired
Emu ( <i>Dromaius novaehollandiae</i> )	genomisch	Gehirn	SRP019803	Illumina HiSeq 2000	68,379,320	80 bp paired
Emu ( <i>Dromaius novaehollandiae</i> )	transkriptomisch	Embryo	SRP019802	Illumina HiSeq 2000	51,118,472	80 bp paired
Emu ( <i>Dromaius novaehollandiae</i> )	transkriptomisch	„generic sample“	SRP001362	LS454	182,922	274 bp single
Weißkehlammer ( <i>Zonotrichia albicollis</i> )	transkriptomisch	Gehirn	SRP029385	Illumina HiSeq 2000	196,980,056	100 bp single
Huhn ( <i>Gallus gallus</i> )	transkriptomisch	Embryo	DRP000595	Illumina HiSeq 2000	461,468,546	100 bp single
Huhn ( <i>Gallus gallus</i> )	transkriptomisch	Gehirn	SRX196389 SRX196371 SRX196380	Illumina HiSeq 2000	54,177,792 235,457,560 64,532,328	35 bp paired 80 bp paired 40 bp paired
Huhn ( <i>Gallus gallus</i> )	transkriptomisch	Testis	SRX196397 SRX196388 SRX196379	Illumina HiSeq 2000	74,846,788 22,373,296 231,583,782	40 bp paired 36 bp paired 80 bp paired
Huhn ( <i>Gallus gallus</i> )	transkriptomisch	Hypothalamus	SRX316899 SRX316900	Illumina HiSeq 2000	191,561,330 188,549,958	100 bp paired
Halsbandschnäpper ( <i>Ficedula albicollis</i> )	transkriptomisch	Embryo	ERX144565 - ERX144572	Illumina Genome Analyzer IIx	49,600,302	100 bp paired
Japanwachtel ( <i>Coturnix japonica</i> )	genomisch	-	DRX001717	Illumina HiSeq 2000	865,334,134	100 bp paired
Zebrafink ( <i>Taeniopygia guttata</i> )	transkriptomisch	Gehirn	SRX738987	Illumina HiSeq 2500	9,535,066	150 bp paired

### 2.5.3 *De novo* Assemblierungen

Die gefilterten Daten wurden mit der Funktion „*de novo assembly*“ der CLC Genomics Workbench (CLC bio, Aarhus, Dänemark) unter Verwendung der Standardparameter assembliert. Der hierbei verwendete Algorithmus nutzt de Bruijn Graphen für die Assemblierung [198]. Die so erhaltenen Contigs wurden dann mit Hilfe von lokalen BLASTn- und BLASTx-Suchen [199] mit *stk33*-Datenbanken, welche alle bekannten *stk33* Nukleotid- und Proteinsequenzen enthielten, auf *stk33*-Contigs untersucht.

### 2.5.4 Kartierungen

Zusätzlich zu den *De Novo* Assemblierungen wurden Kartierungen der gefilterten Rohdaten durchgeführt. Zur Kartierung der transkriptomischen Daten gegen transkriptomische Contigs wurde die Funktion „*map reads to reference*“ und für die Kartierung gegen genomische Contigs die Funktion „*large gap read mapping*“ der CLC Genomics Workbench (CLC bio, Aarhus, Dänemark) verwendet. Für Kartierungen innerhalb einer Spezies wurde eine Ähnlichkeit von mindestens 98 % über 95 % der gesamten Sequenzlänge vorausgesetzt. Für Kartierungen zwischen verschiedenen Arten wurde diese Ähnlichkeit auf 80 % über 80 % der Sequenzlänge reduziert. Hierbei wurden für Fehlpaarungen zwei Strafpunkte und für Insertionen oder Deletionen drei Strafpunkte vergeben.

### 2.5.5 Genomische Analyse der Klasse *Aves*

Zur Analyse der genomischen Struktur von *stk33* in der Klasse *Aves* war es zunächst notwendig, die genomischen Zielregionen der verschiedenen Vogelarten zu identifizieren. Da *stk33* in allen bisher untersuchten Vertebraten stromaufwärts von *lmo1*, *ric3* und *tub* und stromabwärts von *st5*, *rpl27a* und *trim66* liegt, wurde dieser Bereich als annotierte gkb-Datei aus der NCBI UniGene-Datenbank (<http://www.ncbi.nlm.nih.gov/unigene>) heruntergeladen und in die CLC Main Workbench v6.9.1 (CLC bio, Aarhus, Dänemark) importiert. Wenn dieser Bereich nicht vollständig in der UniGene-Datenbank vorhanden war, wurden die Contigs des entsprechenden Genomprojektes als fasta-Datei aus der NCBI Genome Datenbank (<http://www.ncbi.nlm.nih.gov/genome/>) heruntergeladen und in die CLC Main Workbench importiert (vgl. Tabelle 3.6). Die Identifizierung *stk33*-relevanter Contigs erfolgte dann mittels lokaler BLASTn-Suchen [199] unter Verwendung der genomischen *stk33*-Sequenz des Afrikanischen Vogelstraußes (*Struthio camelus*) als Referenz.

Gegen die so identifizierten Contigs erfolgte eine Kartierung der Exon-Sequenzen des Afrikanischen Vogelstraußes (*Struthio camelus*), um so die Exon-Sequenzen in den zu untersuchenden Vögeln identifizieren zu können. Anhand von BLASTx-Suchen [199] der identifizierten Exons mit dem STK33-Protein des Afrikanischen Vogelstraußes (*Struthio camelus*) als Referenz konnten vorzeitige Stopkodons innerhalb der kodierenden Sequenz erkannt werden. Zur Identifizierung von genomischen Deletionen erfolgte ein Alignment der zu untersuchenden genomischen Sequenz gegen die genomische Sequenz des Afrikanischen Vogelstraußes (*Struthio camelus*) mit Hilfe der Funktion „Align two or more sequences“ des NCBI BLASTn [199]. Um abzuklären, ob eine Duplikation des *stk33* Gens abseits des untersuchten Locus existiert, wurden die in Tabelle 3.6 dargestellten Contigs mittels lokaler BLASTn-Suchen mit den genomischen Contigs des Afrikanischen Vogelstraußes (*Struthio camelus*) nach potenziellen *stk33*-Duplikationen durchsucht. Analog hierzu wurde die Möglichkeit einer Retrotransposition abgeklärt. Hierzu wurden BLASTn-Suchen mit dem transkriptomischen Contig des Afrikanischen Vogelstraußes (*Struthio camelus*) durchgeführt.

### 2.5.6 Analyse des Reichs *Animalia* mittels „Data-Mining“

Zur Analyse der genomischen Struktur von *stk33* innerhalb des Reichs *Animalia* (ohne *Vertebrata*), wurden die bereits assemblierten Genome der zu untersuchenden Arten aus der Genomdatenbank des NCBI ([ncbi.nlm.nih.gov/genome](http://ncbi.nlm.nih.gov/genome)) heruntergeladen (Tabelle A 1). Die darin enthaltenen Contigs wurden dann, in Abhängigkeit von Genomgröße und der durchschnittlichen Größe der Gene, mittels des Tools „FastaSlice“ in kleinere Subcontigs mit einer Größe zwischen 4 kb und 10 kb zerlegt, um die anschließenden BLAST-Analysen zu erleichtern. Diese erfolgten zunächst mit Hilfe des BLASTx-Algorithmus und einer selbst erstellen und ständig aktualisierten BLAST-Datenbank, bestehend aus allen bisher gefundenen STK33 Proteinen, um aus dem kompletten Genom alle Contigs mit signifikanten STK33-Treffern ( $E \leq 10^{-5}$ ) herauszufiltern. Die so identifizierten Contigs wurden dann mit Hilfe des BLASTx-Algorithmus gegen die nicht redundante Proteindatenbank des NCBI (nr/nt) weiter analysiert und alle *stk33*-relevanten Contigs extrahiert und annotiert.

Die Analyse auf transkriptomischer Ebene erfolgte sowohl über SRA-Datensätze, als auch über bereits assemblierte Transkriptome aus der TSA-Datenbank des NCBI (<http://www.ncbi.nlm.nih.gov/nucore>). Die Analyse der transkriptomischen SRA-Daten erfolgte wie unter 2.5.1-2.5.4 beschrieben. Die bereits assemblierten Transkriptome wurden mittels BLASTx-Analysen auf *stk33*-relevante Contigs hin untersucht und ggf. annotiert.

## 2.6 Puffer und Lösungen

<b>Agarosegel (1-2,5 %)</b>	1 – 2,5 % (w/v) Agarose in TBE-Puffer (1 x)
<b>Agar-Platten</b>	15 g Agar-Agar 5 ml Ampicillin (10 mg/ml) 0,2 mM IPTG (dioxanfrei) 0,5 % (w/v) X-Gal in 1 ml DMF in 500 ml 1 x LB-Medium
<b>Ampicillin-Stammlösung (100 x)</b>	10 mg/ml Ampicillin in HPLC-H <sub>2</sub> O
<b>Annealing Puffer (1 x)</b>	10 mM Tris, pH7,5 - 8 50 mM NaCl 1 mM EDTA
<b>Base Solution (50 x)</b>	1,25 M NaOH 10 mM EDTA
<b>Bromphenolblau DNA-Ladepuffer</b>	4 M Harnstoff 0,1 M EDTA- Dinatriumsalz 50 % (w/v) Saccharose 0,1 % (w/v) Bromphenolblau
<b>DEPC-H<sub>2</sub>O</b>	1 ml DEPC, ad 1 Liter VE-H <sub>2</sub> O Inkubation über Nacht unter Rühren bei 37 °C 2 x autoklaviert
<b>Dialysepuffer (10 x)</b>	3 M NaCl 0,25 M Tris-HCl 1 M Na <sub>2</sub> EDTA
<b>E-Puffer (10 x)</b>	0,36 M TRIS 0,3 M NaH <sub>2</sub> PO <sub>4</sub> 0,1 M EDTA in VE-Wasser
<b>E3-Medium (60 x)</b>	0,3 M NaCl 10 mM KCl 20 mM CaCl <sub>2</sub> · 2H <sub>2</sub> O 25 mM MgCl <sub>2</sub> · 6H <sub>2</sub> O pH 7,2 (NaOH)
<b>E3-Medium (1 x)</b>	16,5 ml 60 x E3 Medium 100 µl 1 % Methylenblau ad 1 L VE-H <sub>2</sub> O
<b>Ethidiumbromid-Färbelösung</b>	0,001 % (v/v) Ethidiumbromid- Stammlösung in 1x E-Puffer
<b>Ethidiumbromid-Stammlösung</b>	0,5 % (w/v) Ethidiumbromid in 1x E-Puffer

---

<b>Extraktionspuffer (DNA)</b>	10 mM Tris 2 mM EDTA 0.2 % Triton X-100 200 µg/ml Proteinase K
<b>IPTG-Stammlösung</b>	47,6 mg/ml IPTG in DMF
<b>LB-Medium</b>	0,005 % (w/v) Natriumchlorid 0,005 % (w/v) Hefe-Extrakt 0,01 % (w/v) Trypton in VE-H <sub>2</sub> O
<b>MOPS (10 x)</b>	0,2 M MOPS 0,05 M Natriumacetat 0,01 M EDTA pH 7
<b>Neutralisierungslösung</b>	2 M Tris-HCl (pH 5)
<b>Tricaine</b>	0,2 % in H <sub>2</sub> O
<b>Tris-Borat-Elektrophorese-Puffer (1 x TBE)</b>	90 mM Tris 90 mM Borsäure <sup>1</sup> 25 mM EDTA-Dinatriumsalz in VE-H <sub>2</sub> O
<b>X-Gal-Stammlösung</b>	100 mg/ml X-Gal in DMF

# 3 | Ergebnisse

## 3.1 Genome engineering

Zwar konnten aufgrund vorangegangener Untersuchungen bereits einige Informationen über die Lokalisation sowie die Funktion und die Substrate von STK33 gewonnen werden [20; 23; 24], jedoch ist die genaue physiologische Rolle von STK33 weiterhin weitgehend unbekannt. Daher war es ein Ziel dieser Arbeit, einen Zebrafisch „Knock-out“ Stamm für *stk33* zu etablieren, um so die Hypothesen der Rolle von STK33 in Embryogenese und Spermatogenese [20] überprüfen zu können. Hierzu musste zunächst der Zebrafisch als Versuchstier sowie eine entsprechende Anlage etabliert werden. Für das Ausschalten von *stk33* im Zebrafisch fiel die Wahl zunächst auf die zu Beginn dieser Arbeit neu verfügbaren „*Transcription activator-like effector nucleases*“ (TALENs; vgl. 1.3.2) [194].

### 3.1.1 TALEN-vermittelter „Knock-Out“ von *stk33* im Zebrafisch

Bei der Auswahl geeigneter Zielsequenzen mit Hilfe des „*TAL Effector Nucleotide Targeter 2.0*“ der Cornell University konnten im 5'-Bereich der kodierenden Sequenz nur zwei Zielsequenzen ohne zu erwartende „*off-target*“ Effekte identifiziert werden (vgl. Abbildung 3.1 A). Alle anderen Zielsequenzen lagen entweder zu weit im 3'-Bereich des Gens, so dass eine „Restfunktion“ des nicht betroffenen 5'-Bereiches nicht ausgeschlossen werden konnte, oder es wären zahlreiche „*off-target*“ Effekte im Genom des Zebrafisches zu befürchten gewesen. Von diesen beiden Zielsequenzen wurde zunächst jene aus dem dritten kodierenden Exons gewählt und das entsprechende TALEN-Paar nach dem von Sanjana *et al.* publizierten Protokoll [194] hergestellt (vgl. 2.4.1). Dieses basiert auf einer hierarchischen Ligation mittels „*Golden Gate*“ Strategie sowie der anschließenden *in-vitro* Transkription der hergestellten TALEN-Vektoren. Hierbei werden zunächst jeweils sechs der TALEN-Monomere zu zirkulären Hexameren ligiert und anschließend amplifiziert. Je drei dieser dabei entstehenden linearen Hexamere werden dann in einer zweiten „*Golden Gate*“-Reaktion in den TALEN-Zielvektor ligiert und bilden dann



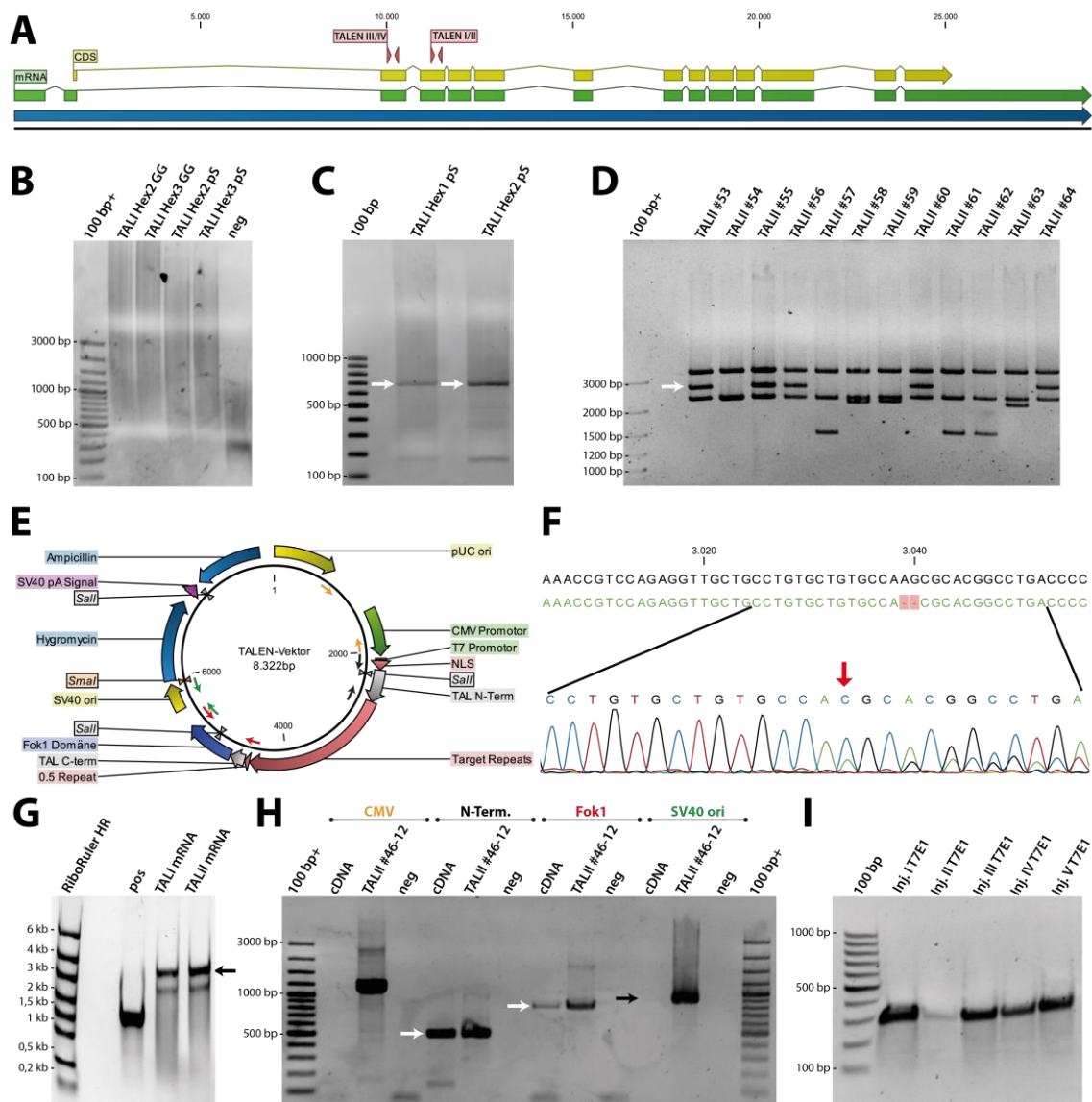
zusammen mit dem auf den Vektor kodierten Monomer für das notwendige Thymin am 5'-Ende sowie dem Halbmer am 3'-Ende der Zielsequenz den vollständigen TALEN-„repeat“ [194]. Aufgrund der Komplexität dieses Verfahrens kam es bereits bei der Herstellung und Transkription der TALEN-Vektoren zu mehreren Problemen. So war nach der erfolgreichen Herstellung der TALEN-Monomer Bibliothek und der ersten „Golden-Gate“ Ligation von je sechs Monomeren zunächst eine spezifische Amplifikation der aus dieser Reaktion entstandenen zirkulären Hexamere nicht möglich. Nach der elektrophoretischen Analyse der durch die Amplifikation entstandenen Produkte zeigte sich nur ein Schmier, ohne die spezifische Bande mit der zu erwartenden Größe von ca. 700 bp zu zeigen. Auch die Verwendung von DMSO oder speziellen Puffern für die Amplifikation GC-reicher DNA sowie verschiedenste Anpassungen der PCR-Parameter inklusive höherer Annealing Temperaturen brachten keine Verbesserung (Daten nicht gezeigt). Auch die Verwendung verschiedener Polymerasen für die Amplifikation führte in den meisten Fällen zu keinem spezifischen Amplifikat (Abbildung 3.1 B). Erst mit Verwendung der Herculase II Polymerase (Agilent, Santa Clara, USA) konnten die zirkulären Hexamere erfolgreich amplifiziert werden (Abbildung 3.1 C). Jedoch war auch hier bei der elektrophoretischen Analyse der Amplifikate im Hintergrund immer noch ein deutlicher Schmier zu erkennen. Außerdem zeigten sich zusätzliche, nicht erwartete Banden mit einer Größe von knapp 200 bp. Nach der Wiedergewinnung der linearen TALEN Hexamere aus dem Agarosegel konnten diese dann mit einer weiteren „Golden Gate“-Reaktion in den jeweiligen Zielvektor kloniert werden.

Die Überprüfung der so assemblierten TALEN-Vektoren auf ihre Vollständigkeit erfolgte zunächst mittels der von Sanjana *et al.* [194] empfohlenen Kolonie-PCR auf zufällig ausgewählte Klone, da mit dem hier verwendete Vektorsystem (vgl. Abbildung 3.1 E) eine Blau-Weiß Selektion nicht möglich ist. Diese Methode führte jedoch, vermutlich aufgrund der Repetitivität der TALEN-Monomere, nicht zu zuverlässigen und aussagekräftigen Ergebnissen. Daher wurden die zufällig ausgewählten Klone zunächst aufgearbeitet und anschließend durch einen *Sall*-Restriktionsverdau auf ihre Vollständigkeit hin untersucht. Hierbei entstehen, bedingt durch den Zielvektor, Restriktionsfragmente mit einer Größe von 3,3 kb und 2,3 kb. Vollständig assemblierte Klone zeigen zusätzlich ein Restriktionsfragment von 2,7 kb (vgl. Abbildung 3.1 D,E), welches den vollständigen TALEN-„repeat“ beinhaltet. Vollständig assemblierte Klone wurden dann zur weiteren Überprüfung sequenziert (2.2.7), um neben der Vollständigkeit auch die korrekte Reihenfolge der eingebauten Monomere zu überprüfen. Hierbei war jedoch eine aussagekräftige Überprüfung der Plasmide nicht möglich, da es innerhalb der Sequenz der TALEN-„repeats“ zu einem scheinbaren n+1 bzw. n+2 Effekt kam, der sich innerhalb des Alignment mit der jeweiligen Referenzsequenz in einer scheinbaren

Deletion äußerte (vgl. Abbildung 3.1 F). Auffällig war hierbei auch, dass es sich bei der zweiten Sequenz, die im Hintergrund des Chromatogramms läuft, um die des vollständigen und korrekten TALEN-Vektors handelt. Weiterhin war nach ca. 150 bp eine Normalisierung des Elektropherogramms zu beobachten (vgl. Abbildung A 4). Auch die Analyse von in *E. coli* SURE-zellen retransformierter Klone zur Verhinderung von Rekombination zwischen den einzelnen Monomeren der TALEN-„repeats“ und eine von der Fa. StarSeq (Mainz, Deutschland) durchgeführte Hot-Start Sequenzierung lieferten identische Ergebnisse. Um überprüfen zu können, ob ein Großteil der TALEN-Vektoren an dieser Stelle während der Kultur dennoch eine Deletion erworben hat, wurde der entsprechende Klon zunächst mit Hilfe von *Sall* restringiert und diese Fragmente anschließend mit Hilfe des T7E1-Assay (vgl. 2.4.4) analysiert. Die elektrophoretische Analyse der hieraus entstandenen Fragmente zeigte jedoch lediglich die durch den *Sall*-Restriktionsverdau entstandenen Fragmente aber keines der im Falle einer tatsächlich vorliegenden Deletion zu erwartenden Spaltprodukte (siehe Abbildung A 3).

Bei der gelelektrophoretischen Analyse (2.3.2.3) der *in vitro* Transkriptions-Produkte (2.3.5) der TALEN-Vektoren war neben dem erwarteten Produkt von ca. 3,8 kb ein weiteres Produkt mit einem Molekulargewicht von etwa 3 kb zu sehen (Abbildung 3.1 G). Um zu überprüfen, ob es sich hierbei um ein unspezifisches oder ein unvollständiges Transkript handelt, wurde dieses Fragment aus dem RNA-Gel mittels Elektroelution wiedergewonnen (vgl. 2.3.3), in cDNA umgeschrieben (2.3.4) und mittels RT-PCR analysiert. Hierzu wurden insgesamt vier verschiedene Primerkombinationen für die Amplifikation unterschiedlicher Bereiche verwendet (Abbildung 3.1 E). Als Negativkontrolle wurde zunächst die Region des CMV-Promotors (gelbe Pfeile) amplifiziert, da diese aufgrund der Linearisierung mittels *SmaI* nicht im Transkript vorhanden sein sollte. Außerdem wurden der n-terminale Bereich (schwarze Pfeile), der c-terminale Bereich mitsamt der FokI-Nuklease (rote Pfeile) sowie die darauffolgende Region bis zur *SmaI*-Schnittstelle (grüne Pfeile) amplifiziert. Die elektrophoretische Analyse dieser Amplifikation zeigte, dass sowohl der N-terminale Bereich als auch der C-terminale Bereich mitsamt der FOK1-Nuklease erfolgreich amplifiziert werden konnten, während dies für die darauf folgende Region bis zur *SmaI*-Schnittstelle um den Replikationsursprung des SV40 nicht möglich war (vgl. Abbildung 3.1 H). Aufgrund dessen kann davon ausgegangen werden, dass es sich höchstwahrscheinlich lediglich um ein unvollständiges und nicht um ein unspezifisches Transkript handelt.

Nach erfolgter Mikroinjektion der TALEN mRNAs (2.4.3) mit unterschiedlichen Konzentrationen und Injektionsvolumina wurden je vier bis fünf Embryonen vereint und genotypisiert. Bei der gelelektrophoretischen Analyse der Produkte des T7E1-Assay waren keines der zu erwartenden Spaltprodukte von ca. 90 bp und ca. 160 bp detektierbar (Abbildung 3.1 I).



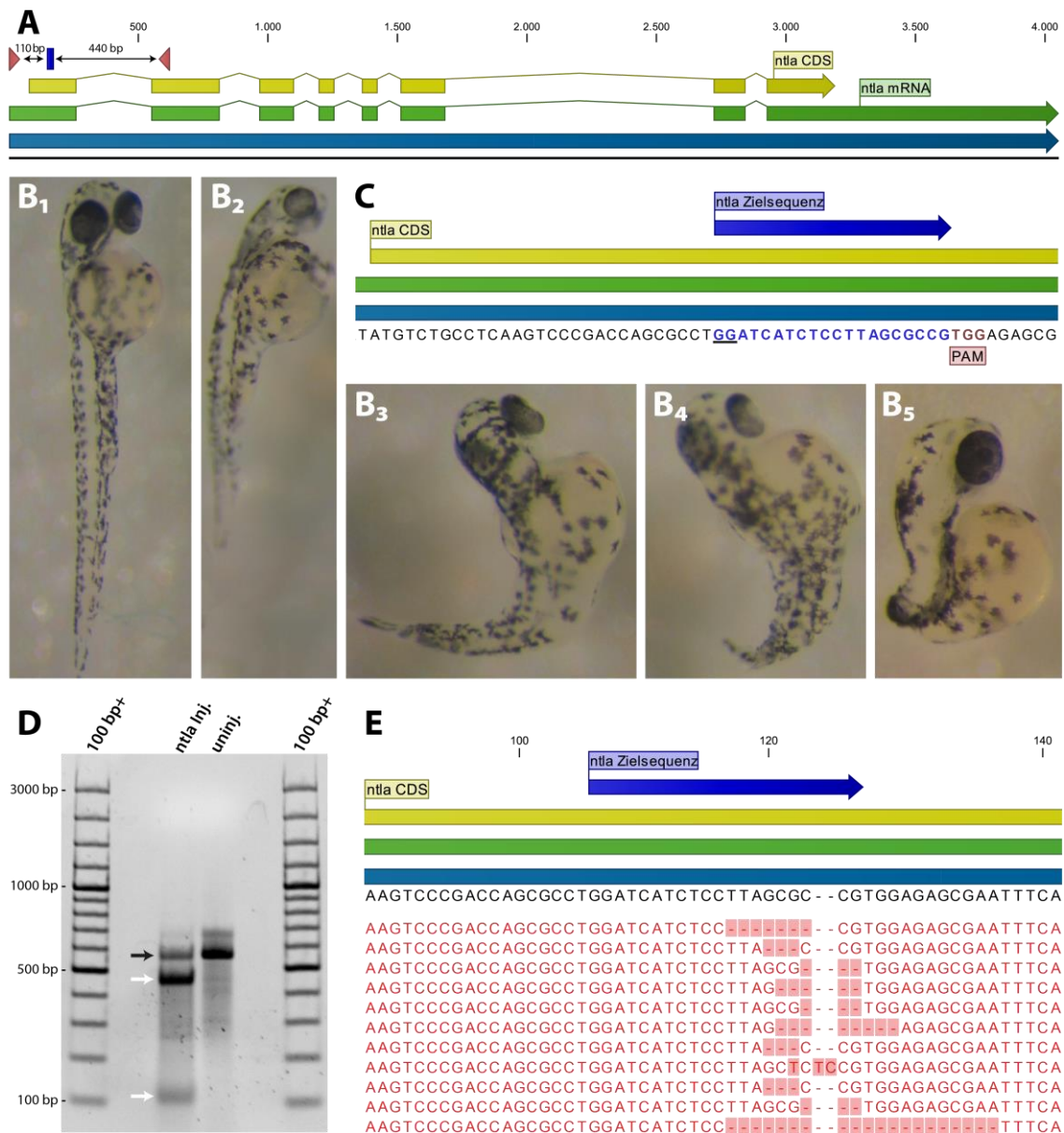
**Abbildung 3.1 | TALEN-vermittelter „Knock-Out“ von *stk33* im Zebrafisch.** (A) Innerhalb des *stk33* Gens im Zebrafisch konnten im zweiten und dritten kodierenden Exon potenzielle Zielsequenzen mit geringen zu erwartenden „off-target“-Effekten identifiziert werden. Alle Introns sind mit 10 % ihrer ursprünglichen Größe dargestellt. (B) Bei der Assemblierung der TALEN-Vektoren war eine spezifische Amplifikation der zirkulären Hexamere zunächst nicht möglich. (C) Erst mit der Verwendung der Herculase II konnte das erwartete Produkt mit einer Größe von ca. 650 bp erhalten werden (weißer Pfeil). (D, E) Zwecks Überprüfung der assemblierten TALEN-Vektoren wurden Restriktionsanalysen mit Hilfe von *SalI* durchgeführt. Hierbei entstehen Restriktionsfragmente mit einer Größe von 3,3 kb und 2,3 kb. Vollständig assemblierte Klone zeigten zusätzlich ein Restriktionsfragment von 2,7 kb (weißer Pfeil). (F) Bei der Sequenzierung vollständiger Klone zwecks Überprüfung der korrekten Reihenfolge der TALEN-Monomere kam es ab bestimmten Positionen (roter Pfeil) zu überlagernden Signalen, welche sich jedoch nach ca. 150 Basen wieder normalisierten. (G) Bei der gelelektrophoretischen Analyse der *in-vitro* Transkription konnte neben der erwarteten Bande von 3,8 kb (schwarzer Pfeil) eine weitere Bande bei ca. 3 kb beobachtet werden. (H) Zur Überprüfung dieses Fragmentes wurde es aus dem Gel wiedergewonnen, in cDNA umgeschrieben und per RT-PCR analysiert. Als Negativkontrolle wurde die Region des CMV-Promotors (gelbe Pfeile) amplifiziert. Zur Analyse des Transkriptes wurden der n-terminale Bereich (schwarze Pfeile), der c-terminale Bereich mitsamt der FokI-Nuklease (rote Pfeile) sowie die darauffolgende Region bis zur *SmaI*-Schnittstelle um den Replikationsursprung des SV40 (grüne Pfeile) amplifiziert. Als Positivkontrolle wurden jeweils die gleichen Bereiche auf dem für die *in-vitro* Transkription verwendeten Klon (TALI #46-12) amplifiziert. Hierbei konnten sowohl der n-terminale Bereich sowie der Bereich um die FOK1-Nuklease erfolgreich amplifiziert werden (weiße Pfeile), während dies für die Region um den Replikationsursprung des SV40 nicht möglich war (schwarzer Pfeil). (I) Bei der Genotypisierung injizierter Embryonen konnten mittels T7E1-Assay keine Mutationen detektiert werden, da neben dem ursprünglichen PCR-Produkt von 258 bp keines der erwarteten Spaltprodukte von ca. 90 bp und ca. 160 bp nachgewiesen werden konnte.

### 3.1.2 CRISPR/Cas-vermittelter „Knock-Out“ von *no-tail (ntl)* im Zebrafisch

Als alternative Methode für das Ausschalten von *stk33* im Zebrafisch wurde das während des Verlaufs dieser Arbeit entwickelte und verfügbar gewordene CRISPR/Cas-System verwendet [53]. Hierzu sollten zunächst „Knock-Out“ Kontrollexperimente mit dem Gen *no-tail (ntl)* durchgeführt werden, um einerseits das Herstellungsverfahren und die Injektion der CRISPR/Cas RNAs zu testen und andererseits die Effizienz und Spezifität des CRISPR/Cas-System zu überprüfen. Der *ntl* Phänotyp zeichnet sich v.a. durch das Fehlen der kaudalen Strukturen aus [200] und stellt somit aufgrund des leicht sichtbaren Phänotyps eine sehr gut geeignete Kontrolle dar.

Als Zielsequenz kam eine bereits verifizierte, 20 bp lange Sequenz (Dr. Holger Dill, IMB, Mainz; persönliche Kommunikation) innerhalb des ersten kodierenden Exons zur Anwendung (siehe Abbildung 3.2 A & C). Die Herstellung der sgRNA und Cas9 mRNA sowie die Mikroinjektion erfolgten wie unter 2.4.2 und 2.4.3 beschrieben. Insgesamt konnten auf diese Weise, verteilt auf mehrere Injektionsrunden, 328 Embryonen injiziert werden. Das Phänotypische Screening der Embryonen erfolgte 48 Stunden nach der Injektion. Hierbei zeigte sich, dass bis zu diesem Zeitpunkt 207 Embryonen (63 %) verstorben waren, während in der uninjizierten Kontrolle nur 10 % - 20 % der Embryonen, je nach Qualität des Geleges, verstorben waren. Von den 121 überlebenden Embryonen konnte bei 78 (65 %) ein eindeutiger *ntl*-Phänotyp in unterschiedlicher Ausprägung beobachtet werden (vgl. Abbildung 3.2 B<sub>2</sub>-B<sub>5</sub>). Dieser reichte von nur leicht verkürzten Schwanzflossen (Abbildung 3.2 B<sub>2</sub>), über deutlich verkürzte Schwanzflossen (Abbildung 3.2 B<sub>3-4</sub>) bis hin zum vollständigen Fehlen der kaudalen Strukturen (Abbildung 3.2 B<sub>5</sub>). Von den verbliebenen 43 überlebenden Embryonen zeigten 22 Embryonen unspezifische Deformationen und 21 Embryonen zeigten keinerlei phänotypische Auffälligkeiten im Vergleich zu den uninjizierten Kontrollembryos (Abbildung 3.2 B<sub>1</sub>).

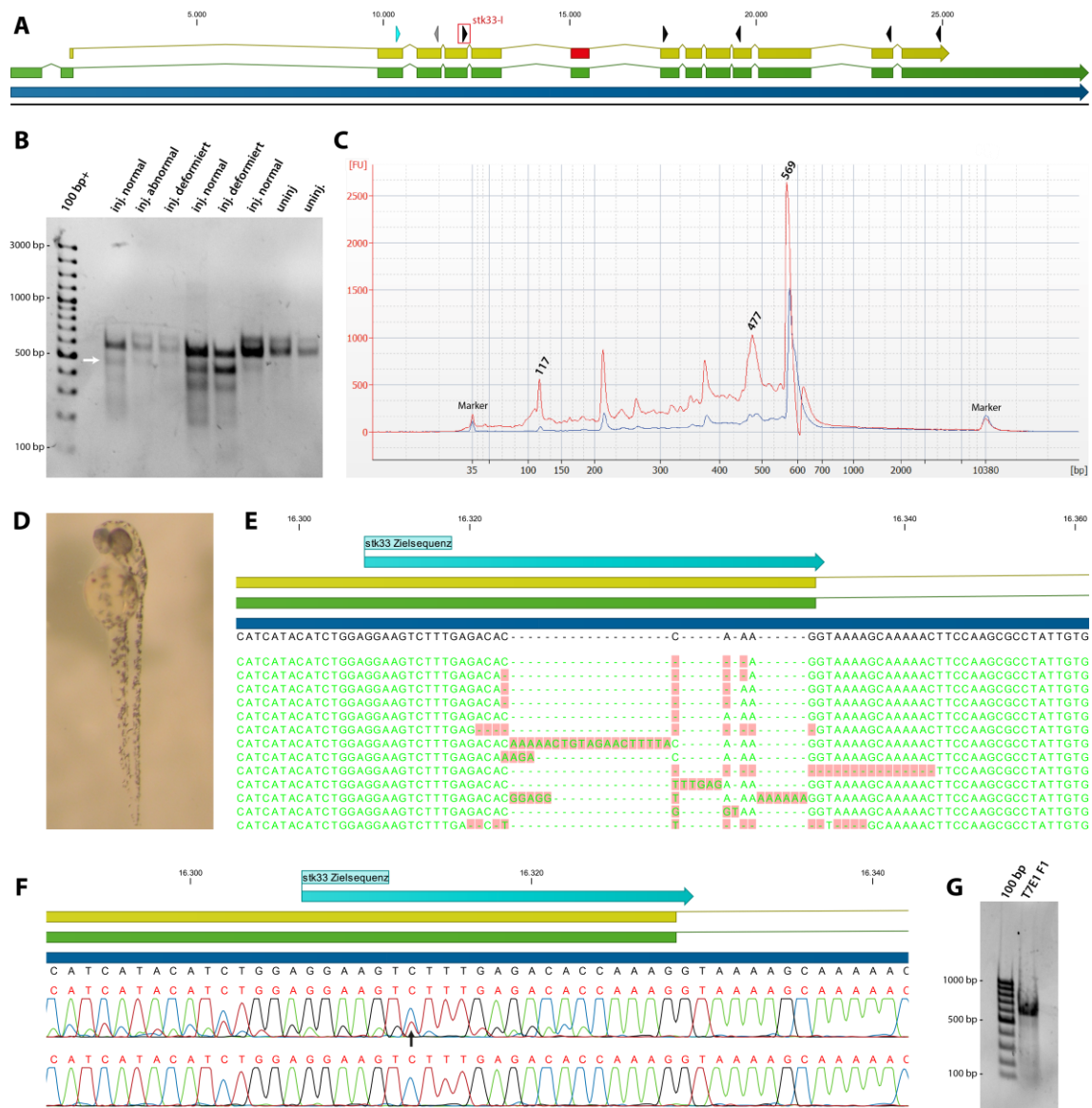
Für die Genotypisierung phänotypisch positiver Embryonen mittels T7E1-Assay (2.4.4) wurden jeweils fünf phänotypisch auffällige Embryonen (vgl. Abbildung 3.2 B<sub>5</sub>) verwendet. Bei der gelelektrophoretischen Analyse der Produkte des T7E1-Assay zeigten sich neben der Bande des ursprünglichen PCR-Produktes (Abbildung 3.2 D, schwarzer Pfeil) deutlich die zu erwartenden Spaltprodukte von 440 bp und 110 bp (Abbildung 3.2 A & D, weiße Pfeile), während diese in der uninjizierten Kontrolle nicht vorhanden waren. Zur genaueren Charakterisierung der genomischen Zielregion wurden die für den T7E1-Assay verwendeten PCR-Produkte kloniert (2.2.6) und anschließend sequenziert (2.2.7). Von den zwölf sequenzierten Klonen enthielten elf Klone DNA Fragmente mit Indels von 2 bp bis zu 18 bp, wobei Deletionen im Vergleich zu Insertionen stark überrepräsentiert waren (Abbildung 3.2 E). Dies entspricht einer Mutationseffizienz in phänotypisch selektierten Embryonen von bis zu 92 %.



**Abbildung 3.2 | CRISPR/Cas-vermittelter „Knock-Out“ von *no-tail* (*ntl*) im Zebrafisch. (A, C)** Die gewählte Zielsequenz (blaues Rechteck) liegt im ersten kodierenden Exon, unmittelbar hinter dem Startkodon. Genannotationen sind in Blau, mRNA Annotationen in Grün und CDS Annotationen in Gelb dargestellt. Lage und Orientierung der Primer für die Genotypisierung sind durch rote Dreiecke gekennzeichnet. Das für die T7 RNA Polymerase notwendige GG-Dinukleotid (unterstrichen) und das „protospacer adjacent motif“ (PAM; in rot) sind gekennzeichnet. Die zu erwartenden Größen der Spaltprodukte des T7E1-Assay sind durch schwarze Pfeile gekennzeichnet. **(B)** 66 % der injizierten Embryonen zeigten im Vergleich zu uninjizierten Kontrollembryos (**B<sub>1</sub>**) den eindeutigen *ntl* Phänotypen in unterschiedlichen Ausprägungen, von nur leicht verkürzten Schanzflossen (**B<sub>2</sub>**), über deutlich verkürzte Schwanzflossen (**B<sub>3,4</sub>**) bis hin zum vollständigen Fehlen der kaudalen Strukturen (**B<sub>5</sub>**). **(D)** Die gelelektrophoretische Analyse der Produkte des T7E1-Assay (2.4.4) zeigte neben der Bande des PCR-Produktes (550 bp; schwarzer Pfeil) deutlich die zu erwartenden Spaltprodukte von 440 bp und 110 bp (weiße Pfeile), während in der uninjizierten Kontrolle diese Spaltprodukte nicht zu sehen sind. **(E)** Bei der genaueren Charakterisierung des genomischen Zielbereiches mittels Klonierung der PCR-Produkte und anschließender Sanger Sequenzierung (vgl. 2.4.4) zeigten sich sowohl Deletionen als auch Insertionen (rote Sequenzen) in elf der zwölf sequenzierten Klone innerhalb des 3'-Bereichs der Zielsequenz, wobei die Deletionen stark überrepräsentiert waren.

### 3.1.3 CRISPR/Cas-vermittelter „Knock-Out“ von *stk33* im Zebrafisch

Auf die erfolgreiche Etablierung des CRISPR/Cas-Systems anhand der *ntl* „Knock-out“ Experimente (siehe 3.1.2) folgten die eigentlichen Versuche zum Ausschalten von *stk33* im Zebrafisch. Hierzu wurde zunächst mittels des webbasierten Programms „ZiFiT Targeter Version 4.2“ des Zinc Finger Consortium [195; 196] der kodierende Bereich von *stk33* des Zebrafisches auf potenzielle Zielsequenzen hin untersucht. Unter Beachtung der - durch die Notwendigkeit für das „*protospacer adjacent motif*“ (NGG) am 3'-Ende der Zielsequenz sowie durch die Transkription der sgRNA mit Hilfe der T7 RNA Polymerase (GG-Dinukleotid am 5'-Ende) - gegebenen Einschränkungen bezüglich der Zielsequenz für die sgRNA (5'-GG-N<sub>18</sub>-NGG-3') [53] konnten in dem 1.503 bp langen kodierenden Bereich von *stk33* lediglich fünf Zielsequenzen identifiziert werden (vgl. Abbildung 3.3 A, schwarze Pfeile). Von diesen fünf Zielsequenzen liegen jedoch vier in den Exons 8, 11, 13 und 14 und somit so weit im 3'-Bereich, dass das durch das Exon 7 kodierte, aktive Zentrum nicht mehr von einer durch das CRISPR/Cas-System induzierten Leserastermutation betroffen wäre. Des Weiteren deutet die Analyse der zu diesen Zielsequenzen gehörenden sgRNAs mit Hilfe des „ZiFiT Targeter“ auf mehrere unspezifische Bindestellen innerhalb des Genoms des Zebrafisches hin. Aufgrund dessen wurde von diesen fünf Zielsequenzen zunächst nur die im vierten kodierenden Exon gelegene Zielsequenz (*stk33-I*; vgl. Abbildung 3.3 A, rotes Rechteck) verwendet, da hier das aktive Zentrum stromabwärts liegt und nur minimale „*off-target*“ Effekte zu befürchten waren. So konnte lediglich eine einzige „*off-target site*“, welche sich nicht innerhalb eines Genes befindet, auf Chromosom 20 an der Position 25.915.248 mb identifiziert werden. Um dennoch weitere Zielsequenzen stromaufwärts des aktiven Zentrums anvisieren zu können, wurden sgRNAs für zwei weitere Zielsequenzen (vgl. Abbildung 3.3 A), die nicht den o.g. Kriterien entsprachen, erstellt. Bei einer dieser Zielsequenzen handelte es sich um eine auf 18 bp verkürzte Zielsequenz (5'-GG-N<sub>16</sub>-NGG-3') für eine sog. „*truncated sgRNA*“ [155]. Die zweite dieser Zielsequenzen beinhaltet nicht das für die Transkription mittels der T7 RNA Polymerase notwendige GG-Dinukleotid am 5'-Ende. Statt dessen ist dieses der Zielsequenz innerhalb der sgRNA vorgeschaltet [146]. Jedoch konnten nach der Injektion dieser sgRNAs keine Mutationen innerhalb der jeweiligen Zielsequenzen festgestellt werden (Daten nicht gezeigt). Die Herstellung der sgRNAs und Cas9 mRNAs sowie deren Mikroinjektion in sich im 1-Zellstadium befindende Embryonen erfolgte wie unter 2.4.2 und 2.4.3 beschrieben. Auf diese Weise konnten, verteilt auf 29 verschiedene Mikroinjektionsrunden, insgesamt 2011 Embryonen für die Zielsequenz *stk33-I* injiziert werden. 48 Stunden nach erfolgter Injektion wurden injizierte sowie uninjizierte Embryonen ausgezählt und entweder als verstorben (†), deformiert (def.) oder normal entwickelt (wildtypisch; wt) klassifiziert (vgl. Tabelle 3.1)



**Abbildung 3.3 | CRISPR/Cas-vermittelter „Knock-Out“ von *stk33* im Zebrafisch.** (A) Übersicht über die Lage der im kodierenden Bereich des *stk33* Locus identifizierten Zielregionen. Genannotationen sind als blaue Pfeile, mRNA-Annotationen als grüne Pfeile und kodierende Bereiche als gelbe Pfeile annotiert. Das für das aktive Zentrum kodierende Exon ist in Rot gezeigt. Alle Introns sind mit 10 % ihrer originalen Größe dargestellt. Alle Zielregionen, welche den Standard-Kriterien des verwendeten CRISPR/Cas-Systems [53] entsprechen (5'-GG-N<sub>18</sub>-NGG-3') sind als schwarze Dreiecke dargestellt, die hiervon verwendete Zielsequenz (*stk33-I*) ist durch ein rotes Rechteck hervorgehoben. Die weiteren, nicht den Standard-Kriterien entsprechenden Zielsequenzen sind als graue (5'-GG-N<sub>16</sub>-NGG-3') bzw. türkisfarbene Dreiecke (5'-NN-N<sub>18</sub>-NGG-3') dargestellt. (B) Das Screening der Embryonen nach erfolgter Injektion mittels T7E1-Assay und die darauffolgende gelelektrophoretische Auftrennung der entstandenen Produkte zeigte deutlich neben der Bande des ursprünglichen PCR-Produktes von 591 bp auch die Bande des zu erwartenden Spaltproduktes von ca. 480 bp. Die Bande des zweiten zu erwartenden Spaltproduktes von ca. 110 bp ist hingegen nicht zu erkennen. In den uninjizierten Kontrollen konnten neben dem PCR-Produkt keine weiteren Produkte nachgewiesen werden. (C) Die Analyse der Spaltprodukte des T7E1-Assay mittels eines „High Sensitivity DNA chip“ auf dem Agilent 2100 Bioanalyzer zeigte neben den Peaks des ursprünglichen PCR-Produktes sowie des Spaltproduktes von ca. 477 bp auch das kleinere Spaltprodukt von ca. 117 bp (rotes Elektropherogramm), während diese Peaks in der uninjizierten Kontrolle nicht zu erkennen sind (blaues Elektropherogramm). (D) Die erfolgreich injizierten Embryonen zeigten, neben unspezifischen Deformationen, keine spezifischen Mutationen. (E) Für genauere Analysen der Zielregion in injizierten Embryonen wurde diese amplifiziert, kloniert und anschließend sequenziert. Hierbei zeigten sich bei 13 der 79 sequenzierten Klone Deletionen oder Insertionen im 3'-Bereich der Zielregion. (F) Nach der Aufzucht erfolgreich injizierter Embryonen wurden diese mit wildtyp-Fischen ausgekreuzt und die daraus resultierende F<sub>1</sub>-Generation auf Ihre Heterozygotie innerhalb der Zielregion hin untersucht. Hierzu wurde zunächst die Zielregion der F<sub>1</sub>-Generation sequenziert. (G) Traten hierbei innerhalb der Zielsequenz Anzeichen für eine mögliche Heterozygotie auf (schwarzer Pfeil), wurde dies mittels eines T7E1-Assay überprüft. Es konnten jedoch keine Mutationen innerhalb der *stk33*-Zielregion identifiziert werden.



**Tabelle 3.1 | Übersicht über die für die Zielsequenz *stk33-I* durchgeführten Mikroinjektionsrunden.** Dargestellt ist die Anzahl der pro Injektionsrunde injizierten Embryonen bzw. der jeweils für die entsprechende uninjizierte Kontrolle verwendeten Embryonen, sowie die Ergebnisse der Auszählung und Bewertung der Embryonen 48 Stunden nach erfolgter Injektion. Hierzu wurden die Embryonen entweder als verstorben (†), deformiert (def.) oder normal entwickelt (wildtypisch; wt) klassifiziert.

Injektionsrunde	Injiziert	†	def.	wt	Uninjizierte Kontrolle	†	def.	wt
1	40	16	14	10	24	7	0	17
2	87	33	30	24	151	23	5	123
3	62	15	31	16	111	9	0	102
4	32	31	1	0	15	13	2	0
5	148	145	0	3	103	85	11	7
6	85	80	5	0	115	40	11	64
7	13	12	0	1	10	10	0	0
8	106	76	20	10	37	32	0	5
9	121	108	5	8	56	40	3	13
10	89	86	0	3	25	10	0	15
11	106	92	0	14	28	9	0	19
12	71	67	0	4	36	18	3	15
13	140	102	0	38	87	28	0	59
14	32	32	0	0	23	6	0	17
15	27	21	0	2	20	6	0	14
16	86	67	8	11	10	6	0	4
17	42	24	6	12	101	31	2	68
18	80	74	5	1	15	10	0	5
19	63	63	0	0	24	14	0	10
20	33	27	3	3	12	8	0	4
21	50	8	20	22	73	11	0	62
22	52	21	0	31	151	36	0	115
23	73	58	6	9	21	8	0	13
24	85	76	0	9	14	5	0	9
25	51	39	7	5	49	7	0	42
26	16	12	0	4	34	5	0	29
27	94	30	15	59	60	19	0	41
28	56	56	0	0	31	31	0	0
29	71	71	0	0	23	23	0	0
<b>Gesamt:</b>	<b>2011</b>	<b>1542</b>	<b>176</b>	<b>299</b>	<b>1459</b>	<b>550</b>	<b>37</b>	<b>872</b>

Von den 2011 injizierten Embryonen waren zu diesem Zeitpunkt insgesamt 1542 (76 % ± 25 %) bereits verstorben und weitere 176 Embryonen (9 % ± 13 %) wiesen 48 Stunden nach erfolgter Injektion deutliche Deformationen auf. Bei diesen Deformationen konnte jedoch kein spezifischer Phänotyp, vergleichbar zu den vorangegangenen *ntl*-Versuchen, identifiziert werden. Lediglich 299 (15 % ± 16 %) Embryonen zeigten bis zu diesem Zeitpunkt eine wildtypische Entwicklung (vgl. Abbildung 3.3 D). Besonders auffällig sind hierbei die relativ hohen Standardabweichungen. Die Überlebensrate der uninjizierten Embryonen liegt typischerweise



zwischen 60 % und 80 %, jedoch kam es auch vor, dass uninjizierte Embryonen eine Überlebensrate von unter 40 % aufwiesen bis hin zu Gelegen, bei denen keine der Embryonen überlebten (vgl. Tabelle 3.1). Aus diesen stark abweichenden Werten ergibt sich eine mittlere Überlebensrate von 50 % mit einer entsprechend hohen Standardabweichung von 29 %. Bei Ausschluss der qualitativ offensichtlich schlechten Gelege (vgl. Tabelle 3.1) ergibt sich eine Überlebensrate von knapp 70 % mit einer deutlich geringeren Standardabweichung von 13 %. Zur Genotypisierung der injizierten Embryonen wurden aus jeder Injektionsrunde, je nach Verfügbarkeit, zwei bis fünf phänotypische normale und ggfs. auch eine gleiche Anzahl an deformierten Embryonen entnommen und mit Hilfe des T7E1-Assay (vgl. 2.4.4) analysiert. Die gelelektrophoretische Analyse der Produkte des T7E1-Assay zeigte deutlich neben der Bande des ursprünglichen PCR-Produktes von 591 bp auch die Bande des zu erwartenden Spaltproduktes von ca. 480 bp (Abbildung 3.3 B, weißer Pfeil). Die Bande des zweiten zu erwartenden Spaltproduktes von ca. 110 bp war hingegen hier nicht zu erkennen. In den uninjizierten Kontrollen konnten neben dem PCR-Produkt keine weiteren Fragmente nachgewiesen werden. Um eine genauere Untersuchung der Spaltprodukte des T7E1-Assay zu ermöglichen, wurden diese mit Hilfe eines „*High Sensitivity DNA chip*“ auf dem Agilent 2100 Bioanalyzer weiter analysiert. Das daraus resultierende Elektropherogramm zeigt deutliche Peaks bei ca. 569 bp und 477 bp. Diese beiden Peaks entsprechen dem ursprünglichen PCR-Produkt und dem größeren der beiden Spaltprodukte. Im Gegensatz zu der gelelektrophoretischen Analyse ist hier auch ein dem zu erwartenden, kleineren Spaltprodukt zuzuordnender Peak bei 117 bp deutlich zu erkennen, während die Analysen der uninjizierten Embryonen diese Peaks nicht zeigen (Abbildung 3.3 C). Zwischen diesen beiden Peaks sind jedoch noch weitere, unspezifische Peaks zu erkennen. Die Untersuchung der potenziellen „*off-target site*“ mittels T7E1-Assay konnte hingegen keine unspezifischen Mutationen der verwendeten sgRNA nachweisen (vgl. Abbildung A 5).

Auch hier sollten, analog zu den vorangegangenen *ntl*-Versuchen (vgl. 3.1.2), die durch das CRISPR/Cas-System in der *stk33-I* Zielregion eingeführten Mutationen näher charakterisiert werden. Hierzu wurde die entsprechende Zielregion amplifiziert, kloniert und anschließend sequenziert. Um eine möglichst hohe Ausbeute an mutanten Sequenzen zu erhalten, wurde die Probe mit der höchsten Mutationsrate für diese Analysen verwendet. Insgesamt konnten 79 Klone sequenziert werden, 13 dieser Klone (16,5 %) enthielten Fragmente mit Insertionen und Deletionen mit einer Länge von 3 bp bis 19 bp im 3'-Bereich der *stk33-I* Zielregion. Auch hier trägt der Großteil der mutanten Fragmente Deletionen, Insertionen scheinen hingegen unterrepräsentiert (siehe Abbildung 3.3 E).

Im Anschluss an die Injektionen wurden alle erfolgreich injizierten und normal entwickelten Embryonen bis zur Pubeszenz (Geschlechtsreife), welche nach drei bis vier Monaten erreicht wird, aufgezogen und anschließend mit einem wildtypischen Partner ausgekreuzt. Aus dieser Auskreuzung sollte dann, sofern die Keimzellen der adulten Tiere eine Mutation innerhalb der *stk33-I* Zielregion tragen, eine heterozygot mutante F<sub>1</sub>-Generation (*stk33<sup>+/-</sup>*) hervorgehen. Aus dieser kann dann durch gezielte Inzucht wiederum eine homozygot mutante F<sub>2</sub>-Generation (*stk33<sup>-/-</sup>*) erzeugt werden (vgl. Abbildung 2.3). Allerdings erreichten lediglich 26 der 299 nach 48 Stunden normal entwickelten Embryonen die Pubeszenz, da der Großteil dieser Tiere noch im Larvenstadium verstarb. Dies entspricht einer Überlebensrate von nur knapp 9 %, während bei einer vergleichbaren Aufzucht wildtypischer Embryonen ca. 80 % bis 90 % der nach 48 hpf („*hours post fertilization*“) vitalen Embryonen die Pubeszenz erreichen.

Für die Analyse der aus der Auskreuzung der geschlechtsreifen Zebrafische hervorgegangenen F<sub>1</sub>-Generation wurden jeweils drei bis fünf Embryonen eines Geleges exemplarisch wie unter 2.4.4 beschrieben mittels Sequenzierung der Zielregion genotypisiert. Zeigte das Chromatogramm hierbei im Bereich der Zielregion Anzeichen für eine mögliche Heterozygotie (vgl. Abbildung 3.3 F), wurde das entsprechende PCR-Produkt zwecks Verifizierung mittels T7E1-Assay weiter analysiert. Jedoch konnten sowohl bei den Sequenzierungen der F<sub>1</sub>-Generation, als auch bei den ggfs. durchgeführten T7E1-Assays keinerlei Mutationen innerhalb der *stk33-I* Zielregion festgestellt werden (vgl. Abbildung 3.3 G).

Um zu überprüfen, ob die adulte F<sub>0</sub>-Generation überhaupt in irgendeiner Form Mutationen für *stk33* aufweist, oder ob alle mutanten Individuen eventuell während des Larvalstadiums verstorben sind und nur nicht mutante Tiere bis zur Pubeszenz heranwachsen konnten, wurden die adulten Tiere im Anschluss an die Genotypisierung der aus der Auskreuzung (vgl. Abbildung 2.3) hervorgegangenen F<sub>1</sub>-Generation vollständig wie unter 2.4.4 beschrieben genotypisiert. Hierbei zeigte sich, dass bei keinem der 26 analysierten adulten F<sub>0</sub>-Tiere Mutationen innerhalb der *stk33-I* Zielregion nachweisbar waren (vgl. Abbildung A 6).

## 3.2 *Stk33* im Reich *Animalia*

Neben der Analyse des Vorkommens der Serin/Threonin Kinase 33 in der Klasse *Aves* (vgl. 3.3) sollte im Rahmen dieser Arbeit auch untersucht werden, ob *stk33* innerhalb des Tierreiches ausschließlich im Unterstamm der Wirbeltiere (*Vertebrata*) vorkommt. Aufgrund der extremen Artenvielfalt, insbesondere innerhalb der Insekten, erfolgten diese Analysen ausschließlich anhand von öffentlich zugänglichen Datensätzen („*Data Mining*“ oder auch „*Knowledge discovery in databases*“). Für die im Rahmen dieser Doktorarbeit durchgeführten Analysen wurde sowohl auf öffentlich verfügbare Genome aus der Genomdatenbank des NCBI als auch auf transkriptomische SRA-Datensätze und bereits assemblierte Transkriptome zurückgegriffen. Auf diesem Wege konnten insgesamt 189 genomische und 537 transkriptomische Datensätze von 488 verschiedenen Tierarten auf das Vorhandensein und die Expression von *stk33* hin untersucht werden.

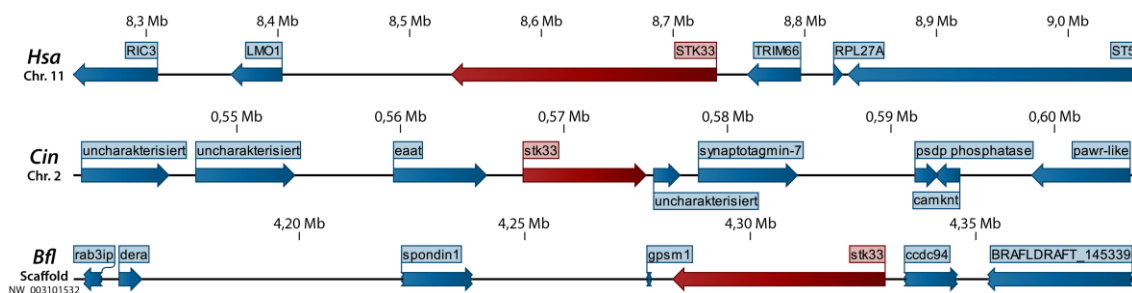
### 3.2.1 *Stk33* in Neumündern (*Deuterostomia*)

Innerhalb des Überstammes *Deuterostomia* wurden zunächst die den Wirbeltieren am nächsten verwandten Taxa innerhalb des Stammes der Chordatiere (*Chordata*) untersucht. Diese schließen, neben den Wirbeltieren (*Vertebrata*), auch die Manteltiere (*Tunicata*, oder auch *Urochordata*) und die Schädellosen (*Cephalochordata*) ein [201]. Mit Bezug auf die innere Systematik der Chordatiere konnten neuere Untersuchungen zeigen, dass nicht die Schädellosen, bestehend aus verschiedenen Gattungen der Lanzettfischchen (*Branchiostoma sp.*), sondern die Manteltiere (*Tunicata*) trotz ihrer stark divergenten Morphologie die nächsten Verwandten der Vertebraten sind [202]. In dem Unterstamm *Tunicata* wurden die Daten der Schlauchseescheide (*Ciona intestinalis*), der Sternseescheide (*Botryllus schlosseri*) und der transparenten Seescheide (*Ciona savignyi*) analysiert. In den Genomen aller untersuchten Arten konnte mittels BLASTx Analysen (2.5.6) eine funktionsfähige Kopie von *stk33* nachgewiesen werden (siehe Tabelle 3.2). Mit Hilfe der verfügbaren transkriptomischen SRA-Daten konnte auch dessen Expression in der Schlauchseescheide und der Sternseescheide nachgewiesen werden (siehe Tabelle 3.2). Für die transparente Seescheide waren keine transkriptomischen Daten verfügbar. In dem Lanzettfischchen *Branchiostoma floridae* (Unterstamm *Cephalochordata*) konnte *stk33* ebenfalls sowohl auf genomischer (Abbildung 3.4), als auch auf transkriptomischer Ebene nachgewiesen werden.

In den Stämmen der Kiemenlochtere (*Hemichordata*) und der Stachelhäuter (*Echinodermata*), zu denen Seegurken, Seesterne, Seeigel und deren Verwandte gehören, wurden insgesamt

sechs verschiedene Arten auf genomischer und transkriptomischer Ebene untersucht. Bei allen konnte die Präsenz und Expression von *stk33* erfolgreich nachgewiesen werden (Tabelle 3.2). Die aus den annotierten Contigs abgeleiteten Proteine zeigen höchst signifikante Ähnlichkeit ( $E=e^{-130}$ ) zu den STK33 Proteinen der Vertebraten. Zwecks der weiterführenden Analysen der gefundenen STK33 Proteine wurden sowohl das webbasierte Programm InterPro (EMBL-EBI; <http://www.ebi.ac.uk/interpro>), als auch Alignments der Proteinsequenzen mit den STK33 Proteinen von Mensch, Maus, Zebrafisch und Vogelstrauß verwendet. Hierbei zeigte sich, dass die in *Deuterostomia* gefundenen Proteine, analog zu den STK33 Proteinen der Wirbeltiere, als Mitglied der Familie der Calcium/Calmodulin-abhängigen Kinasen klassifiziert werden konnten und über die katalytische Domäne der Serin/Threonin-spezifischen Kinasen, eine ATP-Bindestelle sowie das aktive Zentrum der Serin/Threonin-spezifischen Kinasen verfügen (Abbildung 3.5 & Abbildung 3.6).

Bei der Analyse des *stk33* umgebenden Locus zeigt sich, dass die aus den Vertebraten bekannte Syntanie zwischen den Genen *st5* und *ric3* (vgl. Abbildung 1.4 B) in niederen Chordatieren nicht besteht. Hier wird *stk33* von gänzlich verschiedenen Genen flankiert (Abbildung 3.4). Zwar zählen hierzu auch einige noch nicht näher charakterisierte Gene, jedoch weist keines dieser Gene Ähnlichkeiten zu den Genen *st5*, *rpl27a*, *trim66*, *lmo1* oder *ric3* auf. Auch in den unannotierten Bereichen zwischen den einzelnen Genen finden sich keinerlei Hinweise auf die Präsenz dieser Gene. Mit Hilfe genomweiter BLASTx-Suchen konnten im Lanzettfischchen (*Branchiostoma floridae*), mit Ausnahme von *st5*, alle der im Menschen *stk33* flankierenden Gene auf jeweils unterschiedlichen Contigs identifiziert werden. In den *Tunicata* konnten hingegen nur *ric3* und *rpl27a* an unterschiedlichen Positionen auf Chromosom 10 (*ric3* - NC\_020175.2 [365269..369183]; *rpl27a* - NC\_020175.2 [752600..754223]) identifiziert werden.



**Abbildung 3.4 | Der genomische Locus um *stk33* in niederen Chordatieren.** Die flankierenden Gene sind als blaue Pfeile, *stk33* als roter Pfeil dargestellt. Innerhalb der Vertebrata weist der *stk33*-Locus eine hohe Syntanie in dem Bereich zwischen den Genen *st5* und *ric3* auf (vgl. *Homo sapiens* - *Hsa*). In Manteltieren (*Tunicata*; vgl. *Ciona intestinalis* – *Cin*) und Schädellosen (*Cephalochordata*; vgl. *Branchiostoma floridae* – *Bfl*) findet sich diese Syntanie nicht.

### 3.2.2 *Stk33* in Lophotrochozoen (*Lophotrochozoa*)

Die *Lophotrochozoa* bilden zusammen mit den Häutungstieren (*Ecdysozoa*) das Taxon der Urmünder (*Protostomia*). Der Überstamm der *Lophotrochozoa* umfasst neben den Stämmen der Weichtiere (*Mollusca*), Moostierchen (*Bryozoa*) und Armfüßer (*Brachiopoda*) in erster Linie wurmartige Organismen wie die Ringelwürmer (*Annelida*), Spritzwürmer (*Sipuncula*), Schnurwürmer (*Nemertea*), Kelchwürmer (*Entoprocta*) und Hufeisenwürmer (*Phoronidae*) [201]. Die Analyse der Plattwurmartigen (*Platyzoa*), zu denen u.a. die Plattwürmer (*Platyhelminthes*) gehören, soll ebenfalls in diesem Abschnitt behandelt werden, da diese mittlerweile als Subtaxon und nicht als Schwestergruppe angesehen werden [201].

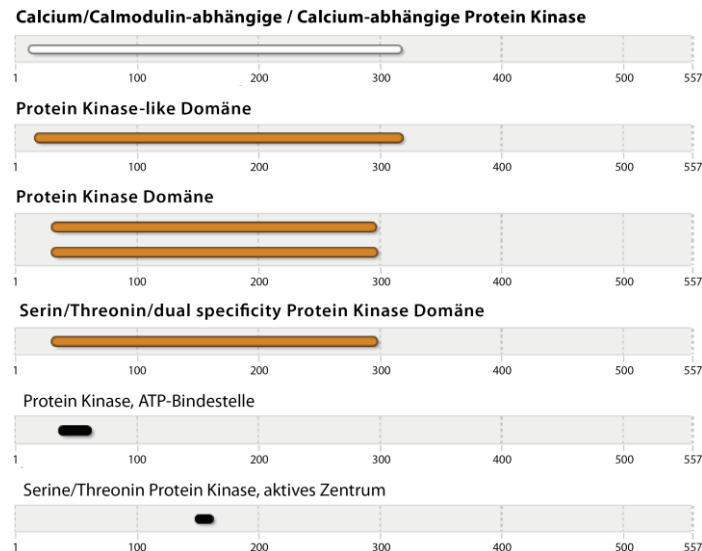
Eine Übersicht über alle untersuchten Arten aus dem Überstamm *Lophotrochozoa* und die Ergebnisse dieser Analysen sind in Tabelle 3.2 zusammengefasst. Insgesamt wurden innerhalb der Lophotrochozoen 43 Tiere analysiert. Hiervon waren für zehn Tiere sowohl genomische als auch transkriptomische Daten aus Geweben vorhanden, in denen eine Expression von *stk33* aufgrund bisheriger Untersuchungen erwartet werden kann [9; 20; 23]. Für 16 der analysierten Organismen standen nur genomische Daten zur Verfügung, während 17 ausschließlich anhand vorhandener transkriptomischer Daten analysiert werden konnten (Tabelle 3.2).

Hierbei konnte neben den großen Stämmen der Weichtiere (*Mollusca*) und Ringelwürmer (*Annelida*) das *stk33* Gen auch in nahezu allen anderen untersuchten Stämmen auf genomischer bzw. transkriptomischer Ebene nachgewiesen werden (vgl. Tabelle 3.2). Eine Ausnahme stellt hier der Stamm der Plattwürmer (*Platyhelminthes*) dar. Innerhalb der Klassen der Hakensaugwürmer (*Monogenea*), Saugwürmer (*Trematoda*) und Strudelwürmer (*Turbellaria*) konnte *stk33* erfolgreich nachgewiesen werden. Innerhalb der Klasse der Bandwürmer (*Cestoda*) war dies für die acht der zur Ordnung *Cyclophyllidea* gehörenden Arten nicht möglich, während für die zwei zur Ordnung *Pseudophyllidea* gehörenden Arten ein Nachweis auf genomischer Ebene geführt werden konnte (siehe Tabelle 3.2).

Auch die aus den in den Lophotrochozoen gefundenen Contigs abgeleiteten Proteinsequenzen sollten mittels Sequenzalignments und InterPro näher charakterisiert werden. Hierbei zeigte sich beim Vergleich zu der Analyse der Daten der Neumünder (siehe 3.2.1) ein identisches Bild. Es handelt sich um Calcium/Calmodulin-abhängige Kinasen, welche über die entsprechende katalytische Domäne, eine ATP-Bindestelle sowie das aktive Zentrum der Serin/Threonin-spezifischen Kinasen verfügen (siehe Abbildung 3.5). Das Alignment der Proteinsequenzen zeigt eine starke Konservierung im Bereich des aktiven Zentrums und eine moderate Konservierung im Bereich der ATP-Bindestelle. Im Bereich der von Mujica identifizierten Asx-reichen Schleife [19] ist hingegen keine Konservierung zu erkennen (vgl. Abbildung 3.6).

**Tabelle 3.2 | Übersicht über untersuchte Arten innerhalb der *Deuterostomia* und *Lophotrochozoa*.** Dargestellt sind die je Taxon analysierten Arten sowie die dazu verwendeten genomischen oder transkriptomischen Datensätze. Ein positiver Befund auf genomischer oder transkriptomischer Ebene ist durch ein (+) in der entsprechenden Spalte gekennzeichnet. War kein Hinweis auf *stk33* in den analysierten Daten vorhanden, so ist dies durch ein (-) gekennzeichnet. Die zur Ordnung Cyclophyllidea gehörenden Bandwürmer (*Cestoda*) sind in grau dargestellt.

Taxon		untersuchte Art	genomisch	transkriptomisch	gen.	trans.	
DEUTEROSTOMIA	Chordata	<i>Tunicata</i>	<i>Ciona intestinalis</i>	PRJDA65419	SRR1324853-72	+	+
			<i>Botryllus schlosseri</i>	PRJNA205369	SRR1604825-62	+	+
			<i>Ciona savignyi</i>	PRJNA1435	x	+	
	<i>Cephalochordata</i>	<i>Branchiostoma floridae</i>	PRJNA33245	SRR923751	+	+	
	<i>Hemichordata</i>	<i>Saccoglossus kowalevskii</i>	PRJNA42857	SRR071700-22	+	+	
	<i>Echinodermata</i>	<i>Ophiothrix spiculata</i>	PRJNA182997	x	+		
		<i>Acanthaster planci</i>	x	SRR1197243		+	
		<i>Patiria miniata</i>	PRJNA49323	SRR573710	+	+	
		<i>Parastichopus parvimensis</i>	PRJNA182998	SRR496203/4	+	+	
		<i>Strongylocentrotus purpuratus</i>	PRJNA13728	SRR532121	+	+	
<i>Mollusca</i>		<i>Aplysia californica</i>	PRJNA209509	GBBG01000000	+	+	
	<i>Lottia gigantea</i>	PRJNA259762	x	+			
	<i>Crassostrea gigas</i>	PRJNA276446	SRR334220/526975	+	+		
	<i>Biomphalaria glabrata</i>	PRJNA290623	SRR942795	+	+		
<i>Annelida</i>	<i>Lumbricus terrestris</i>	x	SRR923752		+		
	<i>Capitella teleta</i>	PRJNA175705	x	+			
	<i>Hermodice carunculata</i>	x	SRR651044		+		
	<i>Hormogaster elisae</i>	x	SRR786597		+		
<i>Nemertea</i>	<i>Lineus longissimus</i>	x	SRR1324981-3		+		
	<i>Tubulanus polymorphus</i>	x	SAMN02739896 SAMN03106206		+		
O	<i>Entoprocta</i>	<i>Loxosoma pectinaricola</i>	x	SRR1611559		-	
		<i>Barentsia gracilis</i>	x	SRR1611554		+	
Z	<i>Phoronidae</i>	<i>Phoronis psammophila</i>	x	SRR1611656		+	
		<i>Phoronis vancouverensis</i>	x	SRR1611666		+	
O	<i>Brachiopoda</i>	<i>Glottidia pyramidata</i>	x	SRR1611555		+	
		<i>Hemithiris psittacea</i>	x	SRR1611556		+	
		<i>Laqueus californicus</i>	x	SRR1611557		+	
		<i>Lingula anatina</i>	PRJNA286275	GDJY01000000	+	+	
C	<i>Gnathostomulida</i>	<i>Austrognathia sp.</i>	x	SRR1976175		+	
		<i>Gnathostomula paradoxa</i>	x	SRR1271607/8/13 SRR1275390		+	
O	<i>Acanthocephala</i>	<i>Macracanthorhynchus hirudinaceus</i>	x	ERR454503/4		-	
		<i>Hymenolepis microstoma</i>	PRJEB124	ERR225717/28-30	-	-	
R	<i>Cestoda</i>	<i>Echinococcus multilocularis</i>	PRJEB122	x	-	-	
		<i>Echinococcus granulosus</i>	PRJEB121	x	-	-	
		<i>Mesocestoides corti</i>	PRJEB510	x	-	-	
		<i>Hymenolepis nana</i>	PRJEB508	x	-	-	
		<i>Hydatigera taeniaeformis</i>	PRJEB534	x	-	-	
		<i>Taenia asiatica</i>	PRJEB532	x	-	-	
		<i>Hymenolepis diminuta</i>	PRJEB507	x	-	-	
		<i>Schistocephalus solidus</i>	PRJEB527	x	+		
		<i>Spirometra erinaceieuropaei</i>	PRJEB1202	x	+		
		<i>Schmidtea mediterranea</i>	PRJNA183840	SRR867386-8	+	+	
T	<i>Turbellaria</i>	<i>Procotyla fluviatilis</i>	x	GAKZ00000000		+	
		<i>Schmidtea polychroa</i>	x	GBZN00000000		+	
		<i>Schistosoma mansoni</i>	PRJEA36577	ERR582645-7/51-53	+	-	
H	<i>Trematoda</i>	<i>Fasciola hepatica</i>	PRJEB6687	ERR576952/3	+	-	
		<i>Echinostoma caproni</i>	PRJEB1207	x	+		
		<i>Dicrocoelium dendriticum</i>	PRJEB3954	x	+		
		<i>Clonorchis sinensis</i>	PRJDA72781	x	+		
		<i>Opisthorchis viverrini</i>	PRJNA259756	PRJNA259756	+	+	
		<i>Schistosoma haematobium</i>	PRJNA273970	PRJNA273970	+	+	
		<i>Trichobilharzia regenti</i>	PRJEB4662	x	+		
<i>Monogenea</i>	<i>Protopolystoma xenopodis</i>	PRJEB1201	x	+			



**Abbildung 3.5 | Analyse der aus den identifizierten mRNA-Sequenzen übersetzten Proteinsequenzen mittels InterPro.** Dargestellt ist, stellvertretend für alle im Reich *Animalia* gefundenen Proteine, die Analyse des STK33-Proteins der Schlauchseescheide (*Ciona intestinalis*). Die Zuordnung des Proteins zu einer Familie ist als weißer Balken dargestellt. Gefundene Domänen sind als orange und funktionale Bereiche als schwarze Balken gezeigt.

### 3.2.3 *Stk33* in Häutungstieren (*Ecdysozoa*)

Die Häutungstiere (*Ecdysozoa*), bilden zusammen mit den *Lophotrochozoa* das Taxon der Urmünder (*Protostomia*) [203]. Die *Ecdysozoa* bestehen neben einigen kleineren Taxa hauptsächlich aus den zwei großen Stämmen der Fadenwürmer (*Nematoda*) und Gliederfüßer (*Arthropoda*) [204]. Die *Arthropoda* stellen dabei mit über einer Million beschriebener Arten den mit Abstand größten Stamm des gesamten Tierreiches dar und repräsentieren somit etwa 60 % aller rezenten Tierarten [205]. Innerhalb der *Arthropoda* entfallen allein knapp eine Million beschriebene Arten auf die Klasse der Insekten (*Insecta*), wobei die Gesamtheit aller rezenten Insektenarten auf sechs bis zehn Millionen geschätzt wird [205–207]. Diese enorme Artenvielfalt ist im Wesentlichen auf die vier „megadiversen“ Ordnungen der Käfer (*Coleoptera*), Schmetterlinge (*Lepidoptera*), Zweiflügler (*Diptera*) und Hautflügler (*Hymenoptera*) zurückzuführen [208; 209].

Neben den großen Stämmen der Fadenwürmer (*Nematoda*) und Gliederfüßer (*Arthropoda*) wurde auch der Stamm der Priapwürmer (*Priapulida*) analysiert. Für den einzigen Vertreter mit verfügbaren Daten, *Priapulius caudatus*, konnte ein funktionales *stk33* sowohl genomisch als auch anhand von transkriptomischen SRA Daten nachgewiesen werden (Tabelle 3.3).

Aus dem Stamm der Fadenwürmer (*Nematoda*) wurden insgesamt 21 Arten aus verschiedenen Klassen und Ordnungen anhand von 20 genomischen und 11 transkriptomischen Datensätzen analysiert. Hierbei konnte in keinem der analysierten Datensätze Hinweise auf *stk33*-relevante Sequenzen gefunden werden (vgl. Tabelle 3.3).

**Tabelle 3.3 | Übersicht über untersuchte Arten innerhalb der Ecdysozoa.** Dargestellt sind die je Taxon analysierten Arten sowie die dazu verwendeten genomischen oder transkriptomischen Datensätze. Ein positiver Fund ist durch ein (+) in der entsprechenden Spalte gekennzeichnet. War kein Hinweis auf *stk33* vorhanden, so ist dies durch ein (-) gekennzeichnet. Die Übersicht über untersuchte Arten aus der Klasse der Insekten ist in Tabelle 3.4 dargestellt.

Taxon	untersuchte Art	genomisch	transkriptomisch	gen.	trans.	
<b>Priapulida</b>	<i>Priapulus caudatus</i>	PRJNA20497	SRR1210404/5 SRR1693484	+	+	
<b>Enoplea</b>	<i>Trichuris muris</i>	PRJEB126	ERR279679/80	-	-	
	<b>Trichocephalida</b>	<i>Trichuris trichiura</i>	PRJEB535	x	-	-
		<i>Trictahuris suis</i>	PRJNA179528	x	-	-
		<i>Trichinella spiralis</i>	PRJNA66457	PRJNA66457	-	-
	<b>Dorylaimida</b>	<i>Longidorus elongatus</i>	x	ERR744622	-	-
	<b>Mermithida</b>	<i>Romanomermis culicivorax</i>	PRJEB1358	x	-	-
	<b>Ascaridida</b>	<i>Ascaris suum</i>	PRJNA80881	x	-	-
	<b>Diplogasterida</b>	<i>Pristionchus pacificus</i>	PRJNA12644	x	-	-
		<i>Subanguina moxae</i>	PRJDB3461	x	-	-
	<b>Tylenchida</b>	<i>Meloidogyne hapla</i>	PRJNA29083	x	-	-
<i>Bursaphelenchus xylophilus</i>		PRJEA64437	ERR864062	-	-	
<b>Spirurida</b>	<i>Onchocerca volvulus</i>	PRJEB513	ERR225734 ERR485008/9/10	-	-	
	<i>Loa loa</i>	PRJNA37757	SRR446605	-	-	
	<i>Brugia malayi</i>	PRJNA27801	PRJNA27801	-	-	
<b>Rhabditida</b>	<i>Steinernema monticolum</i>	PRJNA205067	x	-	-	
	<i>Haemonchus contortus</i>	PRJEB506	SRR928055/6	-	-	
	<i>Panagrellus redivivus</i>	PRJNA186477	SRR678893	-	-	
	<i>Steinernema carpocapsae</i>	PRJNA202318	x	-	-	
	<i>Heterorhabditis bacteriophora</i>	PRJNA13977	x	-	-	
	<i>Caenorhabditis elegans</i>	PRJNA158	PRJNA158	-	-	
<i>Ancylostoma ceylanicum</i>	PRJNA231479	x	-	-		
<b>Symphyla</b>	<i>Symphylella vulgaris</i>	x	GAKX01000000	-	-	
<b>Xiphosura</b>	<i>Limulus polyphemus</i>	PRJNA238073	x	+	-	
	<i>Strigamia maritima</i>	PRJNA20501	x	-	-	
<b>Chilopoda</b>	<i>Scolopendra subspinipes</i>	x	GBIM00000000	-	-	
	<i>Lithobius forficatus</i>	x	GBKE00000000	-	-	
<b>Diplopoda</b>	<i>Glomeris pustulata</i>	x	GAKW00000000	-	-	
	<i>Polydesmus angustus</i>	x	GBKG00000000	-	-	
	<i>Polyxenus lagurus</i>	x	GBKF00000000	-	-	
	<i>Caligus rogercresseyi</i>	PRJNA280098	GAZX00000000	-	-	
<b>Artrhoro-poda</b>	<i>Daphnia pulex</i>	PRJNA12756	x	-	-	
	<i>Eurytemora affinis</i>	PRJNA203087	GBGO00000000	-	-	
	<i>Hyalella azteca</i>	PRJNA243935	GAJQ00000000 GAJP00000000	-	-	
	<i>Lepeophtheirus salmonis</i>	PRJNA40179	x	-	-	
	<i>Triops cancriformis</i>	PRJDB1662	x	-	-	
	<b>Crustacea</b>	<i>Pontastacus leptodactylus</i>	x	GBEI00000000 GAFY00000000 GAFS00000000	-	-
		<i>Calanus finmarchicus</i>	x	GAXK00000000 GBFB00000000 GBXU00000000	-	-
		<i>Eriocheir sinensis</i>	x	GBUF00000000	-	-
		<i>Procambarus clarkii</i>	x	GARH00000000 GBEV00000000	-	-
		<i>Tigriopus californicus</i>	x	GBTC00000000 GBSZ00000000	-	-
<i>Carcinus maenas</i>		x	GBXE00000000	-	-	
<i>Echinogammarus veneris</i>		x	GARO00000000	-	-	
<i>Neomysis awatschensis</i>		x	GDFV00000000	-	-	
<i>Calanus glacialis</i>		x	GBXT01000000	-	-	
<i>Eucyclops serrulatus</i>		x	GARW00000000	-	-	
<i>Tigriopus sp.</i>	x	GDFW00000000	-	-		



Fortsetzung Tabelle 3.3

Taxon	untersuchte Art	genomisch	transkriptomisch	gen.	trans.
<b>Scorpiones</b>	<i>Centruroides exilicauda</i>	PRJNA168116	x	-	
	<i>Mesobuthus martensii</i>	PRJNA171479	x	-	
	<i>Tityus serrulatus</i>	x	GBZU00000000		-
	<i>Latrodectus hesperus</i>	PRJNA168123	GBJN00000000 GBCS00000000	-	-
	<i>Parasteatoda tepidariorum</i>	PRJNA167405	x	-	
<b>Araneae</b>	<i>Pardosa pseudoannulata</i>	x	GCKE00000000 GCVS00000000 GCVT00000000 GCVO00000000		-
	<i>Latrodectus tredecimguttatus</i>	x	GANL00000000		-
	<i>Latrodectus geometricus</i>	x	GBJM00000000		-
	<i>Stegodyphus mimosarum</i>	x	GAZR00000000		-
	<i>Acanthoscurria geniculata</i>	x	GAZS00000000		-
	<i>Sarcoptes scabiei</i>	PRJNA268368	x	-	
	<i>Ixodes ricinus</i>	PRJNA270959	GADI00000000 GCJO00000000 GANP00000000 GBIH00000000	-	-
	<i>Metaseiulus occidentalis</i>	PRJNA166945	x	-	
	<i>Ixodes scapularis</i>	PRJNA34667	PRJNA34667 GBBN00000000	-	-
<b>Acari</b>	<i>Amblyomma americanum</i>	x	GBZX00000000 GAGD00000000		-
	<i>Rhipicephalus annulatus</i>	x	GBJT00000000 GBJS01000000		-
	<i>Amblyomma parvum</i>	x	GBBL00000000		-
	<i>Amblyomma cajennense</i>	x	GBBK00000000		-
	<i>Amblyomma triste</i>	x	GBBM00000000		-
	<i>Ixodes persulcatus</i>	x	GBXQ00000000		-
	<i>Rhipicephalus pulchellus</i>	x	GACK00000000		-
	<i>Ornithodoros turicata</i>	x	GDIE00000000		-
	<i>Dermanyssus gallinae</i>	x	GAIF00000000		-
	<i>Ornithodoros rostratus</i>	x	GCJJ00000000		-
<b>Collembola</b>	<i>Tetradontophora bielensis</i>	x	GAXI00000000		-
	<i>Sminthurus viridis</i>	x	GATZ00000000		-
	<i>Pogonognathellus sp.</i>	x	GATD00000000		-
	<i>Anurida maritima</i>	x	GAUE00000000		+
	<i>Folsomia candida</i>	x	GASX00000000 GAMN00000000		-
	<i>Orchesella cincta</i>	x	GAMM00000000		-
<b>Diplura</b>	<i>Catajapyx aquilonaris</i>	PRJNA203301	x	-	
	<i>Occasjapyx japonicus</i>	x	GAXJ00000000		-
	<i>Campodea augens</i>	x	GAYN01000000		+
<b>Protura</b>	<i>Acerentomon sp.</i>	x	GAXE00000000		-

Die Untersuchung der Gliederfüßer (*Arthropoda*), mit Ausnahme der Insekten, zeigte ein vergleichbares Bild. Die 24 analysierten Spinnentiere (*Arachnida*) und die 17 analysierten Krebstiere (*Crustacea*) sowie einige weitere kleinere Klassen (*Diplopoda*, *Symphyla*, *Chilopoda*) zeigten ebenfalls keine Hinweise auf ein Vorhandensein von *stk33*. Einzig in der Ordnung der Schwertschwänze (*Xiphosura*), vertreten durch den als lebendes Fossil bekannten Atlantischen Pfeilschwanzkrebs (*Limulus polyphemus*) [210], konnte ein Nachweis auf genomischer und transkriptomischer Ebene erfolgreich geführt werden (siehe Tabelle 3.3).

Auch innerhalb der zu den Sechsfüßer gehörenden Klassen der Springschwänze (*Colembola*) und Doppelschwänze (*Diplura*) konnte vereinzelt ein Nachweis für *stk33* auf transkriptomischer Ebene geführt werden, während ein solcher Nachweis in anderen Arten dieser Klassen nicht möglich war (siehe Tabelle 3.3).

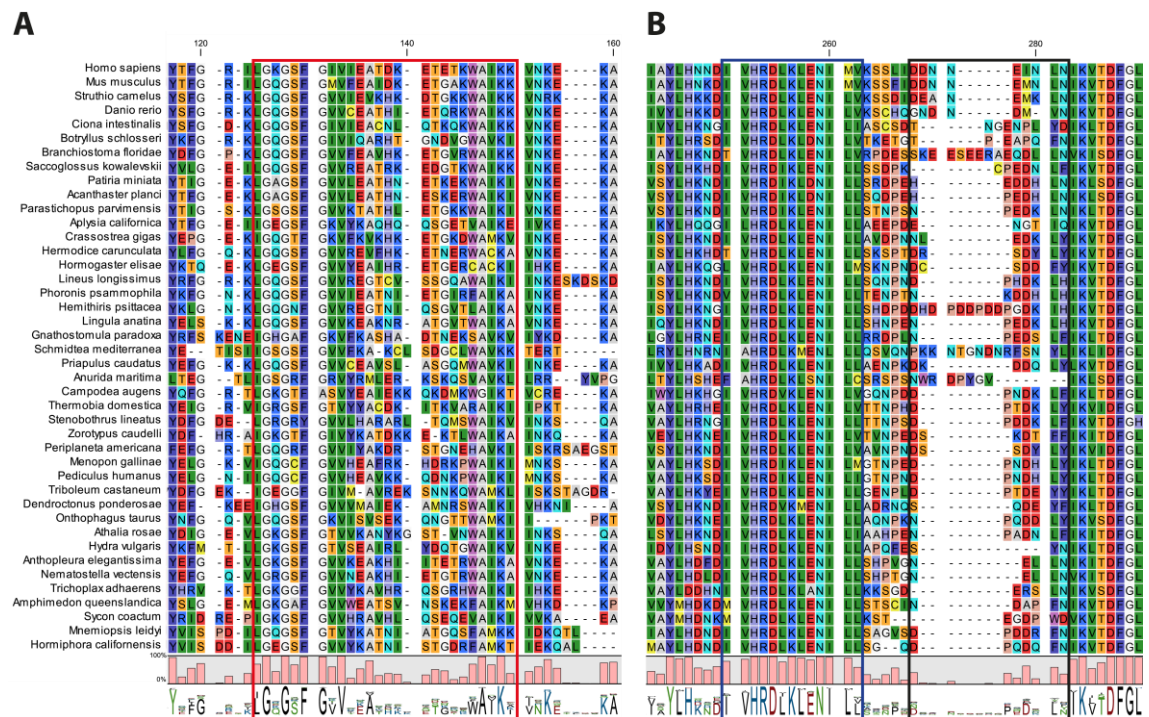
Die Ergebnisse der Analyse der Insekten ist aufgrund der großen Anzahl an analysierten Datensätzen in einer separaten Übersichtstabelle (Tabelle 3.4) dargestellt. Eine detaillierte Übersicht über die verwendeten Datensätze kann Tabelle A 1 (siehe Anhang) entnommen werden. Zwar konnten insgesamt 326 Arten anhand 106 genomischer und 304 transkriptomischer Datensätze innerhalb der Klasse *Insecta* untersucht werden, jedoch stehen diese Datensätze der extremen Artenvielfalt innerhalb der Insekten gegenüber. Dies äußert sich in der geringen Abdeckung einzelner Ordnungen, wie u.a. den Steinfliegen (*Plecoptera*) oder den Flöhen (*Siphonaptera*), insbesondere für den Fall, dass ausschließlich transkriptomische Daten zur Verfügung standen (vgl. Tabelle 3.4). Eine verlässliche Abdeckung mit genomischen und transkriptomischen Datensätzen ergab sich jedoch in den vier „megadiversen“ Ordnungen der Käfer (*Coleoptera*), Schmetterlinge (*Lepidoptera*), Zweiflügler (*Diptera*) und Hautflügler (*Hymenoptera*), sowie in der mit ca. 80.000 bekannten Arten fünftgrößten Ordnung der Schnabelkerfen (*Hemiptera*). Alleine diese fünf Ordnungen repräsentieren über 90 % aller bekannten Insektenarten [208; 209]. In den Ordnungen *Diptera*, *Lepidoptera* und *Hemiptera* konnten keine Hinweise auf *stk33* gefunden werden, während in der Ordnung der Käfer (*Coleoptera*) *stk33* in jedem der sieben analysierten Genome und in 12 der 25 analysierten transkriptomischen Datensätze *stk33* erfolgreich nachgewiesen werden konnte. Eine Besonderheit stellen jedoch die Hautflügler (*Hymenoptera*) dar. Hier wurden insgesamt 30 Genome sowie 52 transkriptomische Datensätze analysiert, jedoch konnte in lediglich zwei Arten (*Fopius arisanus* und *Athalia rosae*) *stk33* nachgewiesen werden, dafür jedoch sowohl auf genomischer und transkriptomischer Ebene (Tabelle 3.4).

Des Weiteren konnte durch die Analyse genomischer und transkriptomischer Daten *stk33*-Contigs auch in den Ordnungen der Heuschrecken (*Orthoptera*), Schaben (*Blattodea*) und Tierläuse (*Phthiraptera*) gefunden werden. Auch für die Ordnungen der Fischchen (*Zygentoma*), Felsenspringer (*Archaeognatha*), Gladiatoren (*Mantophasmatodea*), Bodenläuse (*Zoraptera*), Fangschrecken (*Mantodea*), Kamelhalsfliegen (*Raphidioptera*) und Netzflügler (*Neuroptera*) konnten *stk33*-Contigs auf transkriptomischer Ebene nachgewiesen werden. In den verbleibenden Ordnungen war kein Nachweis von *stk33* möglich (siehe Tabelle 3.4).

Bei der Analyse der gefundenen Proteine zeigte sich, dass auch diese zu der Familie der Calcium/Calmodulin-abhängigen Kinasen gehören und über die entsprechende katalytische Domäne, eine ATP-Bindestelle sowie das aktive Zentrum der Serin/Threonin-spezifischen

Kinasen verfügen (vgl. Abbildung 3.5). Das Alignment mit den bekannten STK33-Proteinsequenzen aus Mensch, Maus, Zebrafisch und Vogelstrauß bestätigt dies. Eine Ausnahme stellt hier das in dem Atlantischen Pfeilschwanzkrebs (*Limulus polyphemus*) gefundene Protein dar. Insgesamt konnten sieben *stk33*-relevante Exons in dem Genom des Pfeilschwanzkrebses identifiziert werden und die Analyse des aus diesen Exons resultierenden Proteins mittels InterPro zeigte ein vergleichbares Ergebnis zu den bisher analysierten Proteinen (Abbildung 3.5), jedoch war das aktive Zentrum der Serin/Threonin spezifischen Kinasen hier nicht nachweisbar. Interessanterweise zeigte sich in dem Bereich zwischen dem dritten und vierten Exon, welche für die das aktive Zentrum flankierenden Aminosäuren kodieren, eine Lücke in der Assemblierung des Genoms.

Das Alignment der im Reich *Animalia* gefundenen STK33-Proteine (Abbildung 3.6) zeigt deutlich, ähnlich wie bereits das Alignment der STK33 Proteine aus der Klasse Aves (vgl. Abbildung 3.9), dass die von Mujica beschriebene Asx-reiche Schleife [19], in unmittelbarer Nachbarschaft zum aktiven Zentrum von STK33, außerhalb der Wirbeltiere nicht mehr konserviert ist (siehe Abbildung 3.6, schwarzes Rechteck).



**Abbildung 3.6 | Alignment ausgewählter, im Reich *Animalia* gefundener STK33-Proteine.** Dargestellt sind ausgewählte Vertreter aus den analysierten Stämmen des Reiches *Animalia*. Als Referenz wurden die STK33-Proteine aus Mensch (*Homo sapiens*), Maus (*Mus musculus*), Afrikanischer Vogelstrauß (*Struthio camelus*) und Zebrafisch (*Danio rerio*) verwendet. (A) Die Region der ATP-Bindestelle ist durch ein rotes Rechteck hervorgehoben, (B) während das aktive Zentrum durch ein blaues Rechteck und die Asx-reiche Schleife durch ein schwarzes Rechteck markiert ist. Der Grad der Konservierung ist durch ein Balkendiagramm verdeutlicht. Im Bereich des aktiven Zentrums zeigt sich eine sehr starke Konservierung der Aminosäuresequenz über alle analysierten Proteine hinweg. Während im Bereich der ATP-Bindestelle noch eine moderate Konservierung zu erkennen ist, ist im Bereich der von Mujica beschriebenen Asx-reichen Schleife [19] außerhalb der Wirbeltiere keine Konservierung mehr festzustellen.

**Tabelle 3.4 | Übersicht über untersuchte Arten innerhalb der Klasse *Insecta*.** Dargestellt sind aus Gründen der Übersichtlichkeit lediglich die Anzahl der je Taxon analysierten Arten sowie die Anzahl der dazu verwendeten genomischen oder transkriptomischen Datensätze. In den beiden letzten Spalten ist angegeben, in wie vielen der verwendeten genomischen bzw. transkriptomischen Datensätzen *stk33*-relevante Contigs nachgewiesen werden konnten. Eine detaillierte Übersicht über die verwendeten Datensätze kann Tabelle A 1 (siehe Anhang) entnommen werden.

	Taxon (Klasse/Ordnung)	# untersuchte Arten	Genomische Datensätze	Transkriptomische Datensätze	gen. +/-	trans. +/-
	<i>Zygentoma</i>	3	0	3		2/3
	<i>Archaeognatha</i>	2	0	2		1/2
	<i>Ephemeroptera</i>	4	1	4	0/1	0/4
	<i>Odonata</i>	5	1	4	0/1	0/4
	<i>Plecoptera</i>	4	0	4		0/4
	<i>Embioptera</i>	2	0	2		0/2
	<i>Phasmida</i>	7	0	7		0/7
	<i>Orthoptera</i>	9	1	10	1/1	2/10
	<i>Mantophasmatodea</i>	1	0	1		1/1
o	<i>Zoraptera</i>	2	0	2		1/2
	<i>Blattodea</i>	4	1	4	1/1	2/4
t	<i>Isoptera</i>	3	1	2	0/1	0/2
	<i>Mantodea</i>	3	0	3		1/3
c	<i>Dermaptera</i>	2	0	3		0/3
	<i>Grylloblattodea</i>	2	0	2		0/2
e	<i>Psocoptera</i>	2	0	2		0/2
	<i>Phthiraptera</i>	2	1	2	1/1	2/2
s	<i>Thysanoptera</i>	5	1	5	0/1	0/5
	<i>Hemiptera</i>	40	12	43	0/12	0/43
n	<i>Megaloptera</i>	3	0	3		0/3
	<i>Raphidioptera</i>	3	0	3		1/3
l	<i>Neuroptera</i>	6	0	6		1/6
	<i>Coleoptera</i>	26	7	25	7/7	12/25
	<i>Strepsiptera</i>	2	1	2	0/1	0/2
	<i>Mecoptera</i>	4	0	4		0/4
	<i>Siphonaptera</i>	4	0	4		0/4
	<i>Trichoptera</i>	6	1	5	0/1	0/5
	<i>Lepidoptera</i>	44	8	51	0/8	0/51
	<i>Hymenoptera</i>	54	30	52	2/30	2/52
	<i>Diptera</i>	72	40	46	0/40	0/46

### 3.2.4 *Stk33* in basalen Tieren

Zu den basalen Vertretern des Reichs Animalia zählen alle nicht bilateral angelegten Tiere, wie die Nesseltiere (*Cnidaria*), die Plattentiere (*Placozoa*) als die strukturell einfachsten vielzelligen Tiere (*Metazoa*) [211], die Schwämme (*Porifera*) sowie die Rippenquallen (*Ctenophora*) [201]. Insgesamt wurden 17 Nesseltiere auf die Präsenz und Expression von *stk33* hin untersucht. Sowohl die Analyse der genomischen Daten, als auch die der transkriptomischen Daten zeigen deutlich, dass *stk33* in diesem Stamm vorhanden ist und auch exprimiert wird. Der Stamm der Plattentiere (*Placozoa*), bestehend aus nur einer einzigen Art, *Trichoplax adhaerens*, dem strukturell einfachsten aller vielzelligen Tiere, verfügt ebenfalls über eine funktionelle Kopie von *stk33* auf genomischer Ebene. Die Transkription von *stk33* in *Trichoplax adhaerens* konnte aufgrund fehlender Daten nicht analysiert werden. Für den Stamm der Schwämme standen

lediglich Genom und Transkriptom von *Amphimedon queenslandica* sowie transkriptomische Daten von zwei weiteren Vertretern zur Verfügung. Dennoch konnten ausnahmslos in allen untersuchten Datensätzen *stk33* Contigs gefunden werden. Für die Rippenquallen (*Ctenophora*) konnte für *Mnemiopsis leidyi* der Nachweis auf genomischer und transkriptomischer Ebene erfolgreich geführt werden, während dies für *Pleurobrachia bachei* nicht möglich war. Bei der Analyse transkriptomischer Daten fünf weiterer Vertreter war ein erfolgreicher Nachweis von *stk33* bei dreien erfolgreich (vgl. Tabelle 3.5).

Auch bei den niederen Tieren zeigte die Analyse der gefundenen Proteine mittels InterPro, dass alle notwendigen Strukturen einer Serin/Threonin Kinase vorhanden sind (vgl. Abbildung 3.5) und wird durch das Alignment dieser Proteine bestätigt (Abbildung 3.6).

**Tabelle 3.5 | Übersicht über untersuchte Arten innerhalb der niederen Tiere.** Dargestellt sind die je Stamm analysierten Arten sowie die dazu verwendeten genomischen oder transkriptomischen Datensätze. Ein positiver Fund auf genomischer oder transkriptomischer Ebene ist durch ein (+) in der entsprechenden Spalte gekennzeichnet. War kein Hinweis auf *stk33* in den analysierten Daten vorhanden, so ist dies durch ein (-) gekennzeichnet.

Stamm	untersuchte Art	genomisch	transkriptomisch	gen.	trans.	
Cnidaria	<i>Hydra vulgaris</i>	PRJNA31231	SRR1024340 GAOL00000000	+	+	
	<i>Pseudodiploria strigosa</i>	PRJEB6871	x	+		
	<i>Thelohanellus kitauei</i>	PRJNA193083	x	+		
	<i>Acropora digitifera</i>	PRJDA67425	x	+		
	<i>Alatina moseri</i>	PRJNA41627	x	-		
	<i>Nematostella vectensis</i>	PRJNA19965	PRJNA19965	+	+	
	<i>Stylophora pistillata</i>	x	GARY00000000		+	
	<i>Anthopleura elegantissima</i>	x	GBYC00000000 GBXJ00000000 SRR1646677		+	
	<i>Aurelia aurita</i>	x	GBRG00000000		+	
	<i>Stylophora pistillata</i>	x	GARY00000000		+	
	<i>Acropora hyacinthus</i>	x	GDIF00000000		-	
	<i>Acropora cervicornis</i>	x	GASU00000000		+	
	<i>Myxobolus cerebrales</i>	x	GBKL00000000		-	
	<i>Hydra oligactis</i>	x	GBFD00000000		-	
	<i>Podocoryna carnea</i>	x	GBEH00000000		+	
	<i>Hydractinia symbiolongicarpus</i>	x	GAWH00000000		+	
	<i>Polypodium hydriforme</i>	x	GBGH00000000		-	
	Placozoa	<i>Trichoplax adhaerens</i>	PRJNA12874	x	+	
	Porifera	<i>Amphimedon queenslandica</i>	PRJNA66531	SRR1511621	+	+
<i>Cliona varians</i>		x	SRR1391011		+	
<i>Sycon coactum</i>		x	SRR504689/90		+	
Ctenophora	<i>Pleurobrachia bachei</i>	PRJNA213480	SRR116669/70 SRR1971491	-	-	
	<i>Mnemiopsis leidyi</i>	PRJNA64405	SRR1971277/84 SRR789900	+	+	
	<i>Pleurobrachia pileus</i>	x	SRR789901		-	
	<i>Hormiphora californensis</i>	x	SRR1992642		+	
	<i>Euplokamis dunlapae</i>	x	SRR777663		+	
	<i>Bolinopsis infundibulum</i>	x	SRR786491		-	
	<i>Beroe abyssicola</i>	x	SRR777787		+	

### 3.3 *Stk33* in der Klasse *Aves*

Innerhalb der Wirbeltiere (*Vertebrata*) konnte die Serin/Threonin Kinase 33 bisher ausschließlich innerhalb der großen Klassen der Säugetiere (*Mammalia*), Reptilien (*Reptilia*), Amphibien (*Amphibia*) sowie in den Knorpel- (*Chondrichthyes*) und Knochenfischen (*Osteichthyes*) nachgewiesen werden (Abbildung A 10). Innerhalb der Klasse der Vögel (*Aves*) konnte solch ein Nachweis bisher nicht geführt werden und es finden sich bereits Hinweise auf ein Fehlen von *stk33* im Huhn (*Gallus gallus*) [212].

Daher war es ein weiteres Ziel dieser Doktorarbeit, die Klasse *Aves* auf das Vorhandensein von *stk33* hin zu untersuchen. Im Rahmen der Diplomarbeit von Steffen Lerch [213] wurde die Analyse des Afrikanischen Vogelstrauß (*Struthio camelus*), einem basalen rezenten Vertreter der Klasse *Aves*, begonnen und im Rahmen dieser Doktorarbeit vervollständigt (vgl. 3.3.1). Die Analyse des Haushuhns (*Gallus gallus domesticus*), als Modellorganismus innerhalb der Klasse *Aves*, wurde im Rahmen der Diplomarbeit von Daniel Schäfer [212] untersucht. Des Weiteren wurden im weiteren Verlauf dieser Doktorarbeit die Transkriptome und Genome von über 30 verschiedenen Vogelarten auf die Präsenz bzw. Expression von *stk33* hin untersucht.

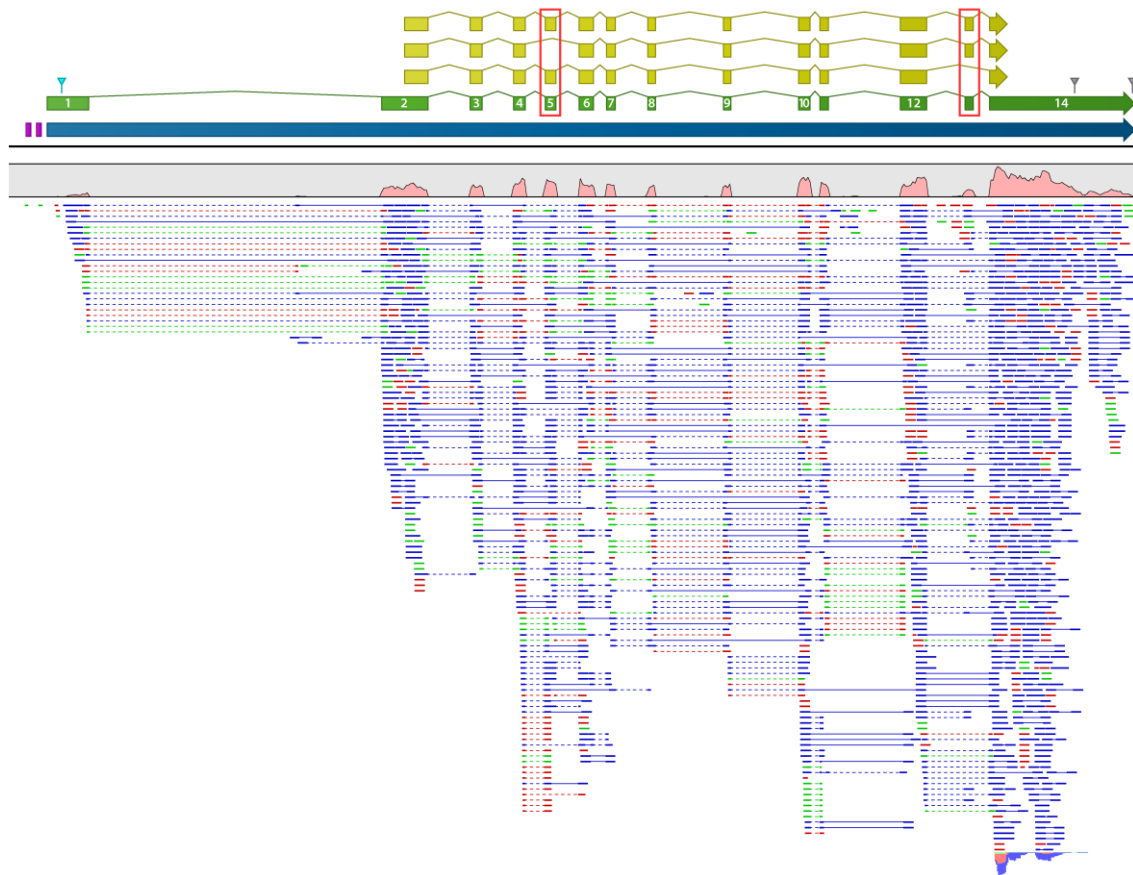
#### 3.3.1 *Stk33* in Urkiefervögeln (*Palaeognathae*)

Zur Aufklärung des genomischen *stk33*-Locus sowie der Intron-Exon Struktur im Afrikanischen Vogelstrauß (*Struthio camelus*) wurde zunächst Total-RNA aus dem Cerebellum und aus Testis adulter Tiere isoliert und zur Erstellung von polyA<sup>+</sup> aufgereinigter cDNA-Bibliotheken verwendet [213], da in diesen Geweben aufgrund der bisherigen Untersuchungen an Mensch und Maus die höchste Expression von *stk33* erwartet werden kann [20]. Die erstellten RNA-Bibliotheken wurden in einem 100 bp „paired-end“ Lauf mittels Illumina® HiSeq2000™-Technologie sequenziert (IMSB, Mainz, Deutschland). Diese Sequenzierung lieferte ca. 10,3 Gigabasen an Sequenzinformationen, welche sich auf 102,624,190 Sequenzen verteilten. Da zu diesem Zeitpunkt keine Referenzsequenz des Genoms zur Verfügung stand, wurden nach dem Qualitätstrimming der beiden Datensätze *de novo* Assemblierungen mit variablen k-mer Größen zwischen 20 und 64 durchgeführt. Diese lieferten zwischen 42.726 und 67.174 Contigs mit einer durchschnittlichen Länge von 781 bp bis 1010 bp. Alle Contigs aus allen Assemblierungen wurden anschließend mit Hilfe von BLASTn Suchen gegen die nicht redundante Nukleotiddatenbank (nr/nt) analysiert. So konnten in 16 der 44 Assemblierungen *stk33*-haltige Contigs identifiziert werden. All diese Contigs wurden erneut assembliert und ergaben einen Scaffold mit einer Länge von 4.020 bp. Um eine vorläufige Annotation der Exons

innerhalb dieses Scaffolds vornehmen zu können, wurde die *stk33* Sequenz des Rotkehlans (*Anolis carolinensis*) [GenBank:XM\_008117649] aufgrund dessen phylogenetischer Nähe zum Vogelstrauß verwendet [213]. Die Überprüfung des transkriptomischen Contigs des Straußes erfolgte mittels RT-PCRs sowie der anschließenden Sanger-Sequenzierung dieser Produkte.

Zur Aufklärung der vollständigen mRNA Sequenz des Vogelstrauß wurden RACE-PCRs durchgeführt (vgl. 2.2.2.3). Hierbei wurden zunächst die Enden des *stk33* Scaffolds mittels BLASTn-Suchen gegen die NCBI Nukleotiddatenbank (nr/nt) analysiert, um zu verhindern, dass die genspezifischen RACE-Primer aufgrund von Fehlern in der Assemblierung außerhalb des tatsächlichen mRNA Moleküls liegen. Hierbei zeigte sich, dass ca. 600 Basen am 5' Ende des Scaffolds signifikante Treffer für das Gen *acvr1* und ca. 800 Basen am 3' Ende Treffer für das Gen *sin3a* ergaben. Diese Bereiche wurden für das Design der genspezifischen RACE-Primer entfernt, da sie höchstwahrscheinlich auf Fehler bei der Assemblierung zurückzuführen waren. Die Ergebnisse der RACE PCRs lieferten sowohl die vollständigen 5'- und 3'-Enden der *stk33* mRNA, als auch einen alternativen Transkriptionsstartpunkt im 5'-Bereich des ersten Exons sowie zwei alternative Polyadenylierungsstellen (siehe Abbildung 3.7). Weiterhin sollte die Intron-Exon-Struktur des *stk33* Gens im Vogelstrauß mit Hilfe intronüberspannender genomischer PCRs und long-range PCRs sowie der anschließenden Sanger-Sequenzierung dieser PCR-Produkte aufgeklärt werden. Diese Untersuchungen lieferten ein 3.366 bp langes *stk33* mRNA Molekül mit insgesamt 14 Exons, die alle von konservierten Spleißstellen (GT/AG) flankiert sind. Aus den Sequenzierungen der genomischen PCRs und Ir-PCRs konnte ein 62.778 bp langer Contig (Abbildung 3.7) erstellt werden [GenBank:KP072780]. In der Promotor-Region konnten zwei GC-Box Elemente an den Positionen -43 und -55 identifiziert werden. Zwecks Verifizierung wurden die transkriptomischen Daten wie unter 2.5.4 beschrieben gegen den genomischen Contig kartiert (Abbildung 3.7).

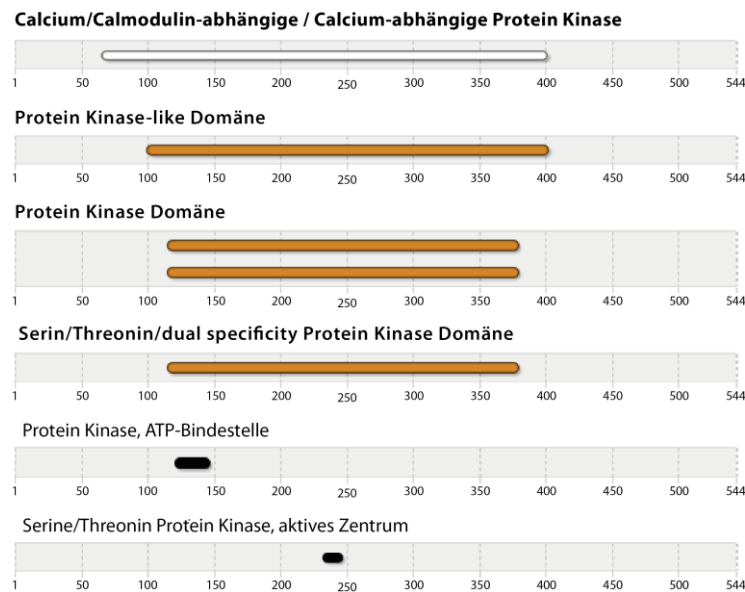
Die *stk33* mRNA Sequenz des Afrikanischen Vogelstraußes enthält einen einzelnen offenen Leserahmen mit einer Länge von 1.635 bp, welcher für ein 544 Aminosäuren langes Protein mit einer Molekülmasse von 59,9 kDa kodiert. Datenbanksuchen mit Hilfe des BLASTp-Algorithmus zeigten, dass dieses Protein signifikante Ähnlichkeit (Identität > 70 %; E=0.0) zu den STK33-Proteinen des Mississippi-Alligators (*Alligator mississippiensis*), der Suppenschildkröte (*Chelonia mydas*), der Chinesischen Weichschildkröte (*Pelodiscus sinensis*) und anderer Reptilien aufweist. Die funktionelle Analyse des gefundenen Proteins des Vogelstraußes erfolgte mittels des webbasierten Programms InterPro. Dieses ermöglicht die funktionelle Analyse von Proteinsequenzen durch die Klassifizierung in Familien sowie die Vorhersage von Domänen und Strukturen durch Vergleiche mit verschiedenen Datenbanken [214].



**Abbildung 3.7 | genomische Struktur des *stk33* Gens im Afrikanischen Vogelstrauß.** Die aus Hoden und Cerebellum gewonnenen transkriptomischen Daten wurden gegen den genomischen Contig des Vogelstraußes kartiert. Gepaarte Reads sind in Blau, verbunden durch eine durchgezogene Linie, forward reads in Grün und reverse reads in Rot dargestellt. Split reads sind durch eine gestrichelte Linie verbunden. Die Sequenzabdeckung ist als roter Graph gezeigt. Das *stk33* Gen ist als blauer Pfeil dargestellt, die dazugehörige mRNA-Sequenz und die jeweiligen kodierenden Sequenzen sind als grüne bzw. gelbe Pfeile dargestellt. Die GC-Boxen sind durch lila Rechtecke dargestellt. Die durch RACE-PCRs identifizierte alternative Transkriptionsstartstelle ist mit einem türkisfarbenen Dreieck markiert, polyA-Signale sind mit grauen Dreiecken markiert. Alle Introns sind aus Gründen der Übersichtlichkeit mit 10 % ihrer originalen Größe dargestellt.

Hierbei konnte das Protein der Familie der Calcium/Calmodulin-abhängigen Kinasen (CAMK) zugeordnet werden. Weiterhin konnte die Serin/Threonin Protein Kinase spezifische katalytische Domäne als funktionale Domäne innerhalb des Proteins identifiziert werden. Als funktionale Sequenzabschnitte innerhalb der katalytischen Domäne konnten eine ATP-Bindestelle sowie das aktive Zentrum einer Serin/Threonin Proteinkinase identifiziert werden (siehe Abbildung 3.8). Die Ergebnisse der Analyse mittels InterPro konnten durch ein Alignment der Proteinsequenz mit der STK33-Sequenz von Mensch, Maus, Zebrafisch und Rotkehlantilope bestätigt werden und es verdeutlicht den hohen Konservierungsgrad innerhalb der Proteinkinase-Domäne, insbesondere im Bereich des aktiven Zentrum (vgl. Abbildung 3.9).





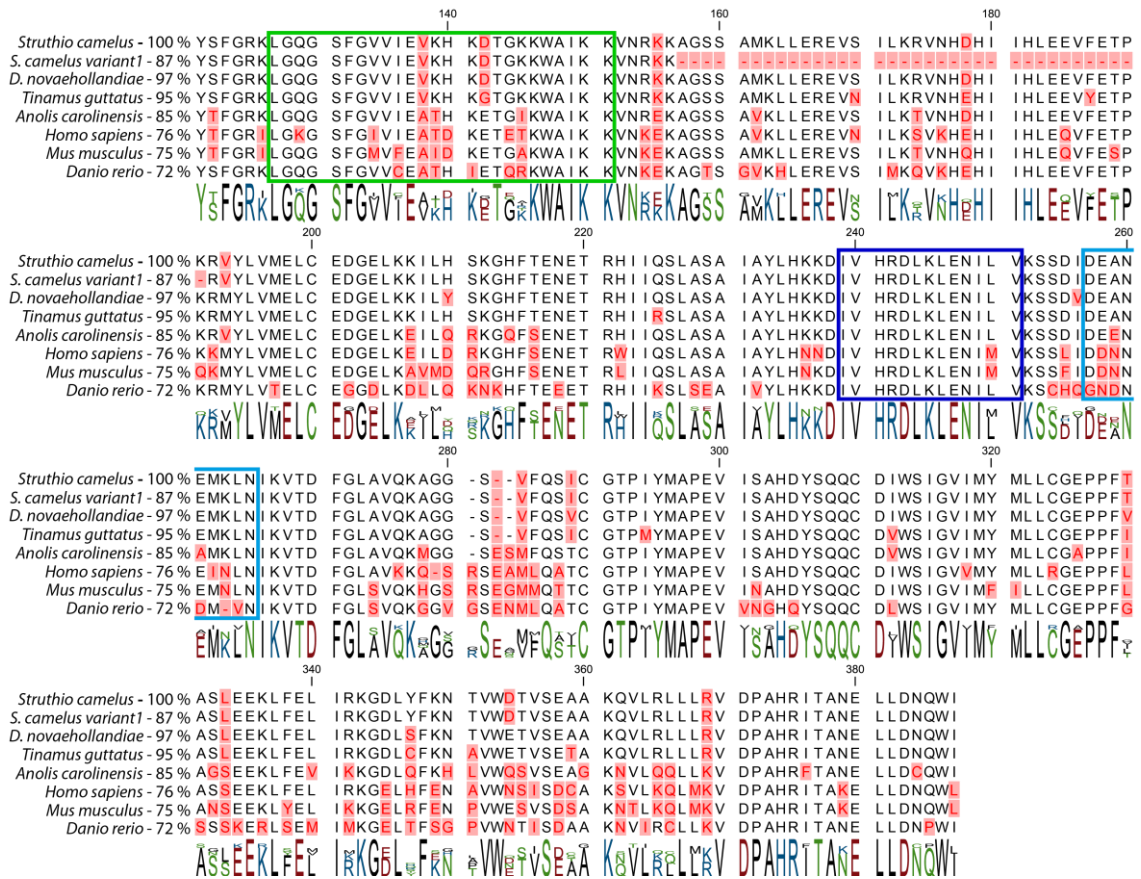
**Abbildung 3.8 | Analyse des gefundenen Proteins des Vogelstraußes mittels InterPro.** Die Übereinstimmung der Proteinssequenz mit den Calcium/calmodulin-abhängigen Kinasen ist als weißer Balken dargestellt. Gefundene Domänen sind als orange Balken und funktionale Sequenzabschnitte sind als schwarze Balken gezeigt. So konnten die gefundenen Proteine der Familie der Calcium/Calmodulin-abhängigen Kinasen zugeordnet werden und es konnte neben der katalytischen Domäne der Serin/Threonin Kinasen auch die ATP-Bindestelle und das aktive Zentrum der Serin/Threonin Kinasen identifiziert werden.

Anhand der Kartierung der transkriptomischen Daten gegen den genomischen Contig konnten Sequenzen identifiziert werden, die sowohl Exon 5 als auch Exon 13 überspannen (Abbildung 3.7). Dies deutet auf das Vorhandensein von alternativ gespleißten Transkripten hin. Um dies zu bestätigen und abzuklären, ob es sich um eine Transkriptvariante, in welcher beide Exons fehlen, oder um zwei verschiedene Transkripte handelt, in denen je entweder Exon 5 oder Exon 13 fehlen, wurden RT-PCRs mit transkriptspezifischen Primern durchgeführt. Hierbei konnte das Vorhandensein von zwei alternativ gespleißten Transkripten nachgewiesen werden. Da Exon 13 außerhalb der katalytischen Domäne liegt, zeigt diese Transkriptvariante hier keinerlei Veränderungen, während die zweite Transkriptvariante durch das Fehlen von Exon 5 eine um 35 Aminosäuren verkürzte katalytische Domäne aufweist. Jedoch zeigten Analysen mittels InterPro, dass weder die ATP-Bindestelle, noch das aktive Zentrum hiervon betroffen sind. Diese Ergebnisse konnten durch ein Alignment der verschiedenen Aminosäuresequenzen bestätigt werden (siehe Abbildung 3.9).

Im folgenden Verlauf sollten die Genome weiterer Urkiefervögel auf das Vorhandensein von *stk33* untersucht werden. Zu diesem Zweck wurden der mit dem Vogelstrauß eng verwandte Große Emu (*Dromaius novaehollandiae*) sowie der Weißkehltinamu (*Tinamus guttatus*) untersucht. Aufgrund der unzureichenden Qualität der öffentlich zugänglichen genomischen SRA-Daten des Emus [SRA:SRP019803] für die Erstellung eines geeigneten genomischen Contigs, wurden ausschließlich transkriptomische SRA-Daten für die Analyse des Emus

verwendet (vgl. Tabelle 2.9). Hierbei konnte eine mRNA Sequenz mit einer Länge von 2.560 bp, bestehend aus 14 Exons, generiert werden. Im Vergleich mit der mRNA des Vogelstraußes zeigt sich eine Identität von 94 % (E=0,0). Diese mRNA enthält einen 1.635 bp langen offenen Leserahmen, der für ein 544 Aminosäuren langes Protein mit einem Molekulargewicht von 60,1 kDa kodiert. Ein Aminosäuren-Alignment zeigt eine Identität von 97 % (E=0,0) innerhalb der katalytischen Domäne (Abbildung 3.9), über die vollständige Sequenz beider Proteine zeigt sich eine Ähnlichkeit von 91 %. Die Analyse des gefundenen Proteins des Großen Emus mittels InterPro zeigte im Vergleich zum Afrikanischen Vogelstrauß ein nahezu identisches Bild (vgl. Abbildung 3.8). Es wurde ebenfalls als Mitglied der Familie der Calcium/Calmodulin-abhängigen Kinasen klassifiziert und enthält weiterhin die Serin/Threonin Protein Kinase spezifische katalytische Domäne, eine ATP-Bindestelle sowie das aktive Zentrum der Serin/Threonin-spezifischen Kinasen. Dies wird auch durch das Alignment mit den STK33 Proteinsequenzen von Mensch, Maus, Zebrafisch und Rotkehlantilope bestätigt (Abbildung 3.9)

Die Untersuchungen des Weißkehlantilopes erfolgten aufgrund des Fehlens von transkriptomischen Daten ausschließlich auf genomischer Ebene. Hierbei zeigt sich, dass das Genom des Weißkehlantilops aufgrund einer großen genomischen Deletion von 2,5 kb, die zum Verlust von Exon 13 führt, lediglich über 13 *stk33*-Exons verfügt [GenBank: BK008887]. Das Fehlen von Exon 13 führt jedoch zu keiner Verschiebung des Leserasters und es finden sich innerhalb des kodierenden Bereiches keine Indels oder Punktmutationen, die zu einer Verschiebung des Leserasters oder zu vorzeitigen Stopkodons führen würden. Dies gilt auch für eine Deletion von 6 bp innerhalb von Exon 12 (Abbildung 3.11). Des Weiteren exprimiert der Strauß eine Spleißvariante ohne Exon 13, welches für die Funktion von STK33 wohl nicht essentiell ist, da es nicht Teil der katalytischen Domäne ist. Durch die Analyse des aus diesem genomischen Contig abgeleiteten Protein mittels InterPro konnte dieses den Calcium/Calmodulin-abhängigen Kinasen zugeordnet und neben der ATP-Bindestelle auch die katalytische Domäne sowie das aktive Zentrum der Serin/Threonin-spezifischen Kinasen identifiziert werden. Auch das Alignment der Proteinsequenzen zeigt eine deutliche Konservierung der ATP-Bindestelle und insbesondere im Bereich des aktiven Zentrums. Außerdem zeigt sich, dass die von Mujica beschriebene Asx-reiche Schleife [19] in den gefundenen STK33-Proteinen aus der Klasse *Aves* nicht konserviert ist (siehe Abbildung 3.9).



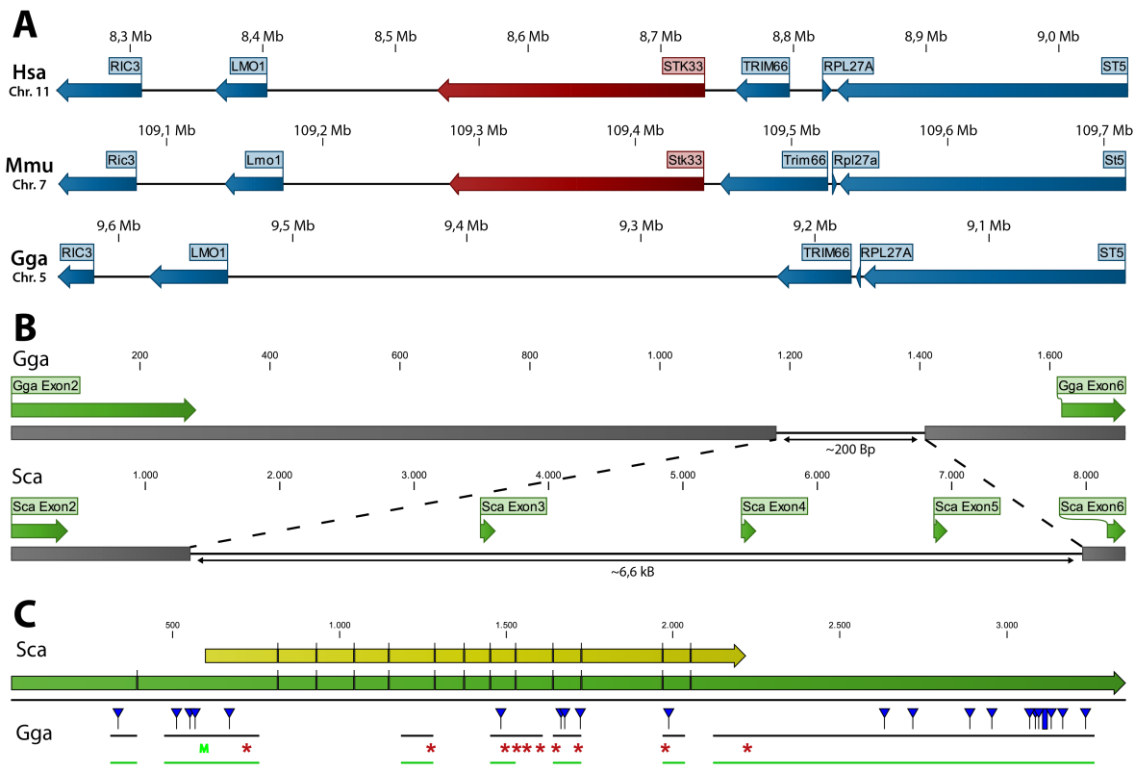
**Abbildung 3.9 | Aminosäuren Alignment katalytischer Domänen verschiedener Spezies.** Dargestellt ist das Alignment der STK33-Proteine von Rotkehlantilope (*Anolis carolinensis*), Mensch (*Homo sapiens*), Maus (*Mus musculus*) und Zebrafisch (*Danio rerio*) als Referenzen sowie den Proteinen des Afrikanischen Vogelstraußes (*Struthio camelus*) als vollständiges Transkript und mit fehlendem Exon 5 (*variant1*), des Emus (*Dromaius novaehollandiae*) und des Weißkehltinamus (*Tinamus guttatus*). Die prozentuale Sequenzidentität innerhalb der katalytischen Domäne ist neben den jeweiligen Sequenzen dargestellt. Die ATP-Bindestelle ist durch ein grünes Rechteck, das aktive Zentrum durch ein blaues Rechteck und die Asx-reiche Schleife durch ein türkisfarbenes Rechteck markiert. Fehlpaarungen sind in Rot hervorgehoben. Deutlich zu erkennen ist die Konservierung der Aminosäuresequenz innerhalb der ATP-Bindestelle und insbesondere im Bereich des aktiven Zentrums, während die Asx-reiche Schleife in den STK33-Proteinen der Vögel nicht konserviert vorliegt.

### 3.3.2 *Stk33* im Haushuhn, *Gallus gallus domesticus*

Die Untersuchung des Haushuhn (*Gallus gallus*) wurde in der Diplomarbeit von Daniel Schäfer [212] begonnen und im Rahmen dieser Doktorarbeit u.a. mit Hilfe der aus der Gruppe der Palaeognathae gewonnen Sequenzinformationen sowie mit der Analyse zusätzlicher transkriptomischer Datensätze vervollständigt.

In Mensch (*Homo sapiens*) und Maus (*Mus musculus*) liegt *stk33* in einem über viele Vertebratenspezies hinweg syntänen Chromosomenabschnitt. Dieser beinhaltet neben *stk33* die konservierten Gene *lmo1* und *ric3* stromabwärts, sowie *trim66*, *rpl27a* und *st5* stromaufwärts von *stk33*. Beim Menschen liegt diese Region auf Chromosom 11p15.3, in der Maus auf Chromosom 7E3. Im Huhn konnte die entsprechende Region auf Chromosom 5 identifiziert werden. Auch hier liegen die Gene *ric3*, *lmo1*, *trim66* und *st5* in konservierter Reihenfolge und Orientierung vor. Das Gen *rpl27a* befindet sich im Genom des Huhns zwar ebenfalls zwischen den Genen *trim66* und *st5*, liegt aber in invertierter Orientierung vor. Im Bereich zwischen den Genen *trim66* und *lmo1*, in dem in Mensch und Maus *stk33* zu finden ist, liegen im Huhn keinerlei Annotationen vor (vgl. Abbildung 3.10 A). Daher stellte dieser Bereich zunächst die Zielregion für die Suche nach *stk33* dar. Die Untersuchung dieser Zielregion erfolgte wie unter 2.5.5 beschrieben. Hierbei konnten zunächst mehrere, größere genomische Deletionen im Vergleich zu der genomischen *stk33*-Sequenz des Vogelstraußes identifiziert werden. Drei dieser genomischen Deletionen haben dabei den vollständigen Verlust von insgesamt sechs Exons (siehe Abbildung 3.10 B) zur Folge. In den verbleibenden genomischen Bereichen konnten die Überreste von acht Exons identifiziert werden. Um Fehler innerhalb der *stk33* Zielregion des Huhns ausschließen zu können, wurde diese Region mittels Sanger-Sequenzierung sowie der Kartierung von SRA-Sequenzen überprüft. Hierbei zeigten sich keine Hinweise auf Fehler innerhalb der genomischen Sequenz [212].

Um eine Zuordnung sowie eine genauere Analyse der verbleibenden Exons zu ermöglichen, wurden diese gegen die *stk33*-mRNA Sequenz des Afrikanischen Vogelstraußes kartiert und der für STK33 kodierende Leserahmen mit Hilfe von BLASTx-Suchen identifiziert. Innerhalb des so identifizierten Leserahmens zeigten sich sowohl zahlreiche durch Indels ausgelöste Leserastermutationen als auch mehrere Stopkodons innerhalb des kodierenden Bereiches, welche zum einen durch die Verschiebung des Leserasters oder zum anderen durch Punktmutationen zustande gekommen sind (vgl. Abbildung 3.10 C). Umfangreiche genomweite BLASTx-Suchen lieferten keinerlei Hinweise auf eine Duplikation oder eine Retrotransposition von *stk33*. Auch die Analyse umfangreicher transkriptomischer Datensätze konnte keine Hinweise auf eine Expression der *stk33*-Exons liefern (siehe 3.3.3). Mit cDNA als Matrize durchgeführte RT-PCRs konnten ebenfalls keine *stk33*-relevanten Amplifikate liefern [212].



**Abbildung 3.10 | *Stk33*-Pseudogen im Haushuhn (*Gallus gallus*).** (A) Die Abfolge und Orientierung der Gene um den genomischen *stk33*-Locus ist zwischen Mensch (Hsa), Maus (Mmu) und Huhn (Gga) konserviert. Im Huhn liegt lediglich *RPL27A* invertiert vor und es finden sich keine Annotationen zwischen den Genen *TRIM66* und *LMO1*. (B) Das Alignment der genomischen *stk33*-Loci des Afrikanischen Volgestraußes (Sca) und des Haushuhns (Gga) zeigen zahlreiche große Deletionen im Bereich des *stk33*-Locus des Huhns, die zum Verlust von insgesamt sechs Exons führen. Exons sind als grüne Pfeile dargestellt und homologe Bereiche sind durch graue Balken gekennzeichnet. (C) Die Kartierung der verbleibenden Exons gegen die mRNA Sequenz des Vogelstraußes (Sca) zeigt zahlreiche Indels und Stoppkodons innerhalb des kodierenden Bereiches. Huhn Exons sind als grüne Linien dargestellt. Die mRNA Sequenz des Vogelstraußes ist als grüner Pfeil, der kodierende Bereich als gelber Pfeil dargestellt. Stoppkodons sind durch rote Sternchen, Startkodons durch den grünen Großbuchstaben „M“ angedeutet. Kleine Indels (<5 bp) sind durch blaue Dreiecke, größere Indels (>5 bp) durch blaue Rechtecke gezeigt.

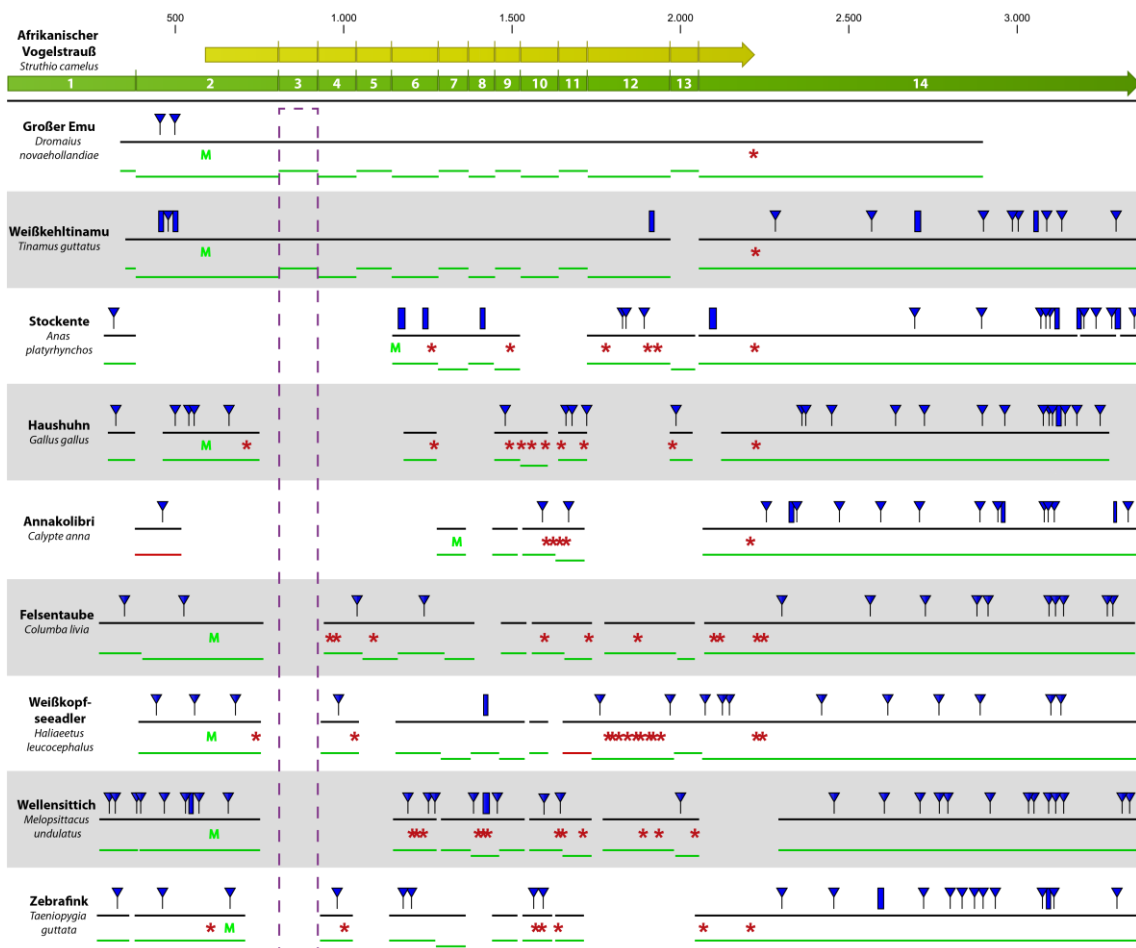
### 3.3.3 *Stk33* in Neukiefervögeln (*Neognathae*)

Um herauszufinden, ob *stk33* in Neukiefervögeln generell fehlt, oder ob das Haushuhn hier eine Ausnahme darstellt, wurden weitere 27 Vertreter aus unterschiedlichen Ordnungen und Familien der Unterklasse Neognathae auf genomischer und vier auf transkriptomischer Ebene mittels bioinformatischer Analyse untersucht (Tabelle 3.6), um eine gute Abdeckung des Stammbaums der Vögel [215–217] zu ermöglichen (vgl. Abbildung A 8) und so eine allgemein gültige Aussage über das Vorhandensein von *stk33* in der Klasse Aves treffen zu können. Die genomischen Untersuchungen erfolgten, analog zu denen des Weißkehltnamus, wie unter 2.5.5 beschrieben.

Bei allen 27 untersuchten Neukiefervögeln konnte der jeweilige *stk33*-Locus identifiziert werden und es zeigte sich ein ähnliches Bild wie im Haushuhn (vgl. 3.3.2). Der genomische *stk33*-Locus ist ebenfalls durch genomische Deletionen in der Größenordnung von

mehreren kb, welche zum Verlust zahlreicher Exons führten, sowie durch Überreste nicht funktioneller *stk33* Sequenzen charakterisiert, während die *stk33* flankierenden Gene intakt zu sein scheinen. Mit Hilfe von BLASTx-Suchen konnten zahlreiche durch Indels verursachte Leserastermutationen und SNPs in den Überresten von *stk33* identifiziert werden, die zur Zerstörung des kodierenden Leserahmens und zur Einführung von vorzeitigen Stopkodons innerhalb des kodierenden Bereiches führten (siehe Abbildung 3.11 & Abbildung A 9).

Selbst die Stockente (*Anas platyrhynchos*), als nächster Verwandter zu den Urkiefervögeln der untersuchten Vögel (siehe Abbildung A 8), verfügt nur noch über ein massiv degeneriertes *stk33*-Gen. Genomische Deletionen führten zum Verlust von insgesamt sechs Exons, durch Indels und SNPs wurden Leserastermutationen und vorzeitige Stopkodons in den noch verbleibenden Exons eingeführt (siehe Abbildung 3.11). Die Genome des Annakolibri (*Calypte*



**Abbildung 3.11 | Alignment identifizierter Exons gegen die *stk33* mRNA Sequenz des Afrikanischen Vogelstrauß.**

Die mRNA Sequenz des Vogelstraußes ist als grüner Pfeil, der offene Leserahmen als gelber Pfeil dargestellt. Grüne Linien kennzeichnen kartierte Exons, rote Linien kennzeichnen invertierte Exons. Kleine Indels (<5 bp) sind durch blaue Dreiecke, größere Indels (>5 bp) durch blaue Rechtecke gezeigt. Stopkodons sind durch rote Sterne und Startkodons durch den grünen Großbuchstaben „M“ angedeutet. Schwarze Linien stellen Konsensussequenzen dar. Die Position von Exon 3 ist durch eine violett gestrichelte Linie hervorgehoben. Während innerhalb der Urkiefervögel keine Deletionen (Ausnahme Exon 13 des Weißkehlтинamu), Leserastermutationen oder vorzeitige Stopkodons auftreten, weist der *stk33*-Locus der Neukiefervögel eine Vielzahl eben solcher Mutationen auf. Allen untersuchten Vogelarten ist die Deletion von Exon 4 gemeinsam.

*anna*) und des Weißkopfseeadlers (*Haliaeetus leucocephalus*) zeigen zusätzlich größere Inversionen innerhalb des *stk33*-Locus, die auch zur Invertierung einzelner Exons führen. Aus der Ordnung der Sperlingsvögel (*Passeriformes*), welche mit knapp 6.000 Arten mehr als die Hälfte aller rezenten Arten der Klasse *Aves* enthält [216], wurden insgesamt sieben Arten untersucht. Hierbei konnte die Deletion der Exons 3, 5, 8, 12 und 13, sowie zahlreiche Leserasterverschiebende Indels und Stopkodons festgestellt werden. Auch alle anderen untersuchten Neukiefervögel zeigen vergleichbare Mutationen innerhalb des *stk33*-Locus. Interessanterweise konnte in allen untersuchten Neukiefervögeln die Deletion von Exon 3 als einzige gemeinsame Mutation nachgewiesen werden (Abbildung 3.11 & Abbildung A 9).

Im weiteren Verlauf sollte zum einen untersucht werden, ob *stk33*, ungeachtet der zahlreichen Mutationen, innerhalb des *stk33*-Locus der Neukiefervögel dennoch exprimiert wird. Zum anderen sollte die Möglichkeit, dass *stk33* eventuell neue Exons erworben hat und so eine funktionelle Kopie des Gens erhalten geblieben ist, untersucht werden. Hierzu wurden verschiedene transkriptomische NGS-Datensätze von Geweben, in denen aufgrund von vorhergegangenen Studien eine Expression von *stk33* zu erwarten war [20; 23] aus dem „Sequence Read Archive“ (SRA; NCBI) auf *stk33*-relevante Sequenzen hin analysiert. Diese Datensätze beinhalteten über 1,5 Milliarden Sequenzen aus Testes, Gehirn und Embryo des Haushuhns (*Gallus gallus*), ungefähr 200 Millionen Sequenzen aus der Weißkehlammer (*Zonotrichia albicollis*) und 50 Millionen Sequenzen aus dem Halsbandschnäpper (*Ficedula albicollis*). Zusätzlich konnten durch eine Sequenzierung von cDNA aus dem Gehirn des Zebrafinken (*Taeniopygia guttata*) auf dem Illumina HiSeq 2500 (IMSB, Mainz, Deutschland) 9,5 Millionen Sequenzen gewonnen werden (siehe Tabelle 2.9) [218]. Diese Datensätze wurden wie unter 2.5.1 und 2.5.2 beschrieben umgewandelt, gefiltert und anschließend sowohl gegen die entsprechende genomische Zielregion, als auch gegen die *stk33*-mRNA Sequenz des Vogelstraußes kartiert (vgl. 2.5.4). Bei diesen Kartierungen konnten keinerlei *stk33*-relevante Sequenzen identifiziert werden. Des Weiteren lieferten RT-PCRs auf die auf genomische Ebene identifizierten Exons keinerlei spezifische Amplifikate [212; 218].

Um auch eine Duplikation oder eine mögliche Retrotransposition im Verlauf der Evolution von *stk33* ausschließen zu können, wurden zum einen transkriptomische *de novo* Assemblierungen durchgeführt (2.5.3) und zum anderen wurden die vollständigen Genome der zu untersuchenden Vogelarten (vgl. Tabelle 3.6) durch umfangreiche BLASTx-Suchen weiter analysiert. Bei diesen genomischen Analysen konnten keine weiteren *stk33*-spezifischen Contigs nachgewiesen werden. Die aus der DeNovo Assemblierung erhaltenen Contigs wurden ebenfalls mittels BLASTx-Suchen untersucht, jedoch konnten auch hier keine *stk33* Contigs gefunden werden.

**Tabelle 3.6 | Übersicht über untersuchte Arten innerhalb der Klasse Aves.** Dargestellt sind die analysierten Arten sowie die dazu verwendeten genomischen oder transkriptomischen Datensätze. Eine funktionsfähige *stk33* Sequenz auf genomischer oder transkriptomischer Ebene ist durch ein (+) in der entsprechenden Spalte gekennzeichnet. War kein Hinweis auf *stk33* in den analysierten Daten vorhanden, so ist dies durch ein (-) gekennzeichnet.

untersuchte Art		gen. Daten	trans. Daten	gen.	trans	Accession no.
Vogelstrauß	<i>Struthio camelus</i>	gen. PCRs lr-PCRs	PRJNA264356 RT-PCR	+	+	KP072780
Emu	<i>Dromaius novaehollandiae</i>	SRP019803	PRJNA193353 SRP001362		+	-
Weißkehltnamu	<i>Tinamus guttatus</i>	PRJNA212876	x	+		BK008887
Annakolibri	<i>Calypte anna</i>	PRJNA212866	x	-		BK009112
Haushuhn	<i>Gallus gallus</i>	PRJNA13342	PRJDB90 SRX196371/79/80/88/89/97 SRX316899/900	-	-	BK009113
Felsentaube	<i>Columba livia</i>	PRJNA167554	x	-		BK009114-16
Sonnenralle	<i>Eurypygia helias</i>	PRJNA212893	x	-		BK009117-18
Truthuhn	<i>Meleagris gallopavo</i>	PRJNA42129	x	-		BK009119-20
Stockente	<i>Anas platyrhynchos</i>	PRJNA46621	x	-		BK009121-23
Seidenreiher	<i>Egretta garzetta</i>	PRJNA232959	x	-		BK009124-26
Sakerfalke	<i>Falco cherrug</i>	PRJNA168071	x	-		BK009127-29
Wanderfalke	<i>Falco peregrinus</i>	PRJNA159791	x	-		BK009130-31
Wellensittich	<i>Melopsittacus undulatus</i>	PRJNA72527	x	-		BK009132
Weißkehlammer	<i>Zonotrichia albicollis</i>	PRJNA197293	PRJNA217789	-	-	BK009133
Tibetmeise	<i>Pseudopodoces humilis</i>	PRJNA175930	x	-		BK009134
Mittel-Grundfink	<i>Geospiza fortis</i>	PRJNA156703	x	-		BK009135-37
Amerikanerkrähe	<i>Corvus brachyrhynchos</i>	PRJNA212869	x	-		BK009138-41
Keilschwanz-Regenpfeiffer	<i>Charadrius vociferus</i>	PRJNA212867	x	-		BK009142-43
Braunflügel-Mausvogel	<i>Colius striatus</i>	PRJNA212892	x	-		BK009144-46
Rotfußseriema	<i>Cariama cristata</i>	PRJNA212889	x	-		BK009147-48
Weißkopfseeadler	<i>Haliaeetus leucocephalus</i>	PRJNA237821	x	-		BK009149
Bergtrogon	<i>Apaloderma vittatum</i>	PRJNA212878	x	-		BK009150
Carolinanachtschwalbe	<i>Caprimulgus carolinensis</i>	PRJNA212888	x	-		BK009151-54
Weißschwanz-Tropikvogel	<i>Phaethon lepturus</i>	PRJNA212902	x	-		BK009155-56
Goldbandpipra	<i>Manacus vitellinus</i>	PRJNA212872	x	-		BK009157-58
Sterntaucher	<i>Gavia stellata</i>	PRJNA212895	x	-		BK009159-60
Japanwachtel	<i>Coturnix japonica</i>	PRJDB1146 DRX001717	x	-		-
Zebrafink	<i>Taeniopygia guttata</i>	PRJNA17289	SRX738987	-	-	BK009161-62
Eissturmvogel	<i>Fulmarus glacialis</i>	PRJNA212894	x	-		BK009163-67
Kea	<i>Nestor notabilis</i>	PRJNA212900	x	-		BK009168-71
Halsbandschnäpper	<i>Ficedula albicollis</i>	PRJNA75089	ERX144565/72	-	-	BK009172



# 4 | Diskussion

## 4.1 Genome Engineering

Um genauere Informationen über die physiologische Rolle von STK33 zu erhalten, war es ein Ziel dieser Doktorarbeit, einen Knock-out Stamm für *stk33* im Zebrafisch (*Danio rerio*) herzustellen. Hierfür kam, nach der erfolgreichen Etablierung eines geeigneten Haltungssystems für den Zebrafisch, zunächst die zu Beginn dieser Arbeit gerade neu etablierte Technologie der „*Transcription activator-like effector nucleases*“ (TALENs; vgl. 1.3.2) [194] zur Anwendung. Im späteren Verlauf wurde diese Technologie dann durch ein speziell für den Zebrafisch entworfenes Protokoll [53] des neu verfügbar gewordenen CRISPR/Cas-Systems ersetzt (vgl. 1.3.3).

### 4.1.1 TALEN-vermittelte „*Knock-Out*“ Versuche

Bei der Auswahl geeigneter Zielsequenzen innerhalb des *stk33* Locus des Zebrafisches wurden die von Sanjana *et al.* postulierten Kriterien [194] angewandt. Bei Berücksichtigung der Notwendigkeit eines Thymins am 5'-Ende der Zielsequenz und einer Gesamtlänge der Zielsequenz von insgesamt 20 bp ergibt sich hieraus eine Zielsequenz der Form 5'-TN<sub>19</sub>-3' für ein einzelnes TALEN-Protein. Da die in den TALENs verwendeten Fok1-Nukleasen aber immer nur als Dimer in der Lage sind, einen Doppelstrangbruch in die DNA einzuführen, muss ein zweites TALEN-Protein auf dem gegenüberliegenden Strang in antiparalleler Orientierung und in einem Abstand von 14 bp bis 20 bp auf der DNA binden. Hieraus ergibt sich dann die entsprechende Zielsequenz der Form 5'-TN<sub>19</sub>N<sub>14-20</sub>N<sub>19</sub>A-3' [194]. Solch eine Zielsequenz findet sich zwar häufig innerhalb des *stk33* Locus, jedoch wurden hier nur solche im 5'-Bereich des durch Exon 7 kodierten aktiven Zentrums berücksichtigt. Nur so kann sicher gewährleistet werden, dass das nach einer erfolgreichen Mutation der anvisierten Zielsequenz gebildete Protein keine Restfunktion mehr aufweist. Dadurch ist die Suche nach einer möglichen Zielsequenz auf den 3'-Bereich von Exon 2 sowie die Exons 3 bis 6 beschränkt (vgl. Abbildung 3.1 A). Weiterhin deutet die Analyse der in diesem Bereich identifizierten Zielsequenzen auf

zahlreiche und gravierende „*off-target*“ Effekte innerhalb des Zebrafischgenoms bei dem Großteil solcher Zielsequenzen hin. Solche unspezifischen Effekte können, je nach Lage, zu der Deaktivierung von regulatorischen Elementen oder anderen Genen [219–222; 154] sowie zu unerwünschten genomischen Rearrangements [45–47] führen. Bei einer allzu großen Anzahl an solchen „*off-target*“-Effekten können diese für die jeweiligen Zellen sogar toxisch sein [93]. Deshalb konnten lediglich zwei Zielsequenzen in dem 5'-Bereich von Exon 7 ohne zu erwartende unspezifische Effekte identifiziert werden. Von diesen wurde zunächst das entsprechende TALEN-Paar (TALEN I/II) für die Zielsequenz im dritten kodierenden Exon (vgl. Abbildung 3.1 A) wie unter 2.4.1 beschrieben hergestellt.

Jedoch kam es bereits bei der Herstellung der TALENs zu einigen Schwierigkeiten. So war zunächst die spezifische Amplifikation der zirkulären Hexamere nach der ersten „*Golden Gate*“ Reaktion auch mit verschiedenen Polymerasen und Reaktionsbedingungen nicht möglich (Abbildung 3.1 B). Diese Problematik ist mit hoher Wahrscheinlichkeit auf die repetitive Struktur der 20 aufeinanderfolgenden TALEN-Monomere zurückzuführen, da die Amplifikation von repetitiver DNA bekanntlich problematisch ist [194; 223; 224]. Erst die Amplifikation mit Hilfe der Herculase II Polymerase (Agilent, Santa Clara, USA) führte zu dem erwarteten Produkt. Bei den nach wie vor vorhandenen unspezifischen Fragmenten handelt es sich wahrscheinlich um unvollständige Transkripte. Allerdings lassen sich hierdurch nicht die gegenüber dem erwarteten Fragment von ca. 700 bp doch deutlich größeren Fragmente sowie die zusätzlich beobachteten Banden unterhalb des erwarteten Produktes erklären. Aufgrund nahezu identischer Probleme bei der Amplifikation der TALEN-Monomere gibt die von Hommelsheim *et al.* postulierte Hypothese hierfür eine mögliche Erklärung. Diese besagt, dass es aufgrund der repetitiven Struktur in einigen Fällen zu einer frühzeitigen Dissoziation der Polymerase von dem DNA-Strang oder durch ein fehlerhaftes Annealing zweier Template-Stränge zu einer Haarnadel-ähnlichen Struktur und somit zum Abbruch der Amplifikation kommen kann. All dies führt zu einer nur unvollständigen Amplifikation der TALEN-Monomere, wobei die daraus entstanden unvollständigen und einzelsträngigen DNA-Moleküle wiederum als sog. „Megaprimer“ oder als neue Templates dienen können. Da diese einzelne oder mehrere der TALEN Monomere enthalten, können diese wiederum mit diesen Monomeren an ähnlichen Stellen innerhalb des TALEN-repeats binden und dort als Primer dienen [224]. Mit Hilfe der Gelextraktion konnten diese unspezifischen Produkte jedoch ausgeschlossen werden. Da bei der Sequenzierung der fertig assemblierten TALEN-Vektoren immer wieder scheinbare Deletionen zu beobachten waren (vgl. Abbildung 3.1 F), der T7E1-Assay jedoch keine Hinweise auf eine Deletion liefern konnte (vgl. Abbildung A 3), kann hieraus geschlossen werden, dass es sich hierbei lediglich um Sequenzierungsfehler, vermutlich aufgrund der Repetitivität des

TALEN-„*repeats*“, handelt. Dies wird insbesondere durch die Beobachtung bestätigt, dass nach ca. 150 bp wieder eine Normalisierung des Elektropherogramms auftritt (siehe Abbildung A 4). Bei einer tatsächlich vorliegenden Deletion wäre dies nicht zu erwarten gewesen.

Eine Möglichkeit zur Vermeidung der hier aufgetretenen Probleme bei der Herstellung der TALEN-Vektoren könnten alternative Herstellungsverfahren zu der hier verwendeten „*Golden Gate*“ basierten Methode sein. Diese alternativen Herstellungsverfahren lassen sich in zwei verschiedene Klassen unterteilen. Die erste dieser beiden Klassen umfasst zwei Methoden („*Unit Assembly*“ und „*REAL – Restriction Enzyme and Ligation*“) zur Herstellung von TALENs durch die sequenzielle Fertigung der TALEN-Vektoren aus vorgefertigten Modulen, jedoch sind hier aufgrund der Modularität enorm viele verschiedene vorgefertigte Plasmide notwendig [182; 225]. Die zweite Klasse von Herstellungsverfahren umfasst mit der FLASH- („*Fast Ligation-based Automatable Solid-phase High-throughput*“) und ICA-Methodik („*Iterative Capped Assembly*“) Techniken zur Hochdurchsatz-Herstellung von TALENs an einer festen Oberfläche. Diese Verfahren ermöglichen zwar die Herstellung von TALENs im Hochdurchsatzverfahren, setzen aber spezielle Ausstattung und Reagenzien voraus [112; 114] und konnten somit im Rahmen der vorliegenden Arbeit nicht durchgeführt werden.

Bei der Genotypisierung der injizierten Embryonen mittels T7E1-Assay zeigten sich keine der bei einer erfolgreichen Mutation der Zielsequenz erwarteten Spaltprodukte (vgl. Abbildung 3.1 I). Anhand dieser Ergebnisse lässt sich feststellen, dass das TALEN-Paar TALEN I/II an der vorgesehenen Zielsequenz entweder keine Mutationen eingeführt hat, oder dass die Mutationsfrequenz unterhalb der Detektionsgrenze von 1 % mutanten DNA-Molekülen [226] des T7E1 Assay liegt. Eine genauere Möglichkeit zur Detektion von Mutationen mit einer Frequenz von unter 1 % bietet die Sequenzierung der Zielregion mit Hilfe der „*Next-Generation Sequencing*“ Technologie, die in der Lage ist, Mutationen mit einer deutlich geringeren Frequenz von 0,1 % bis 0,01 % nachzuweisen [153]. Hiervon wurde im Rahmen dieser Arbeit jedoch aufgrund des hohen Aufwandes und der vergleichsweise hohen Kosten abgesehen.

Eine mögliche Erklärung für diese Ergebnisse ist eine eventuell vorliegende Methylierung innerhalb der Zielregion. Untersuchungen konnten zeigen, dass TALENs nicht in der Lage sind, an methylierte DNA Abschnitte zu binden und dort Doppelstrangbrüche einzuführen [227]. Aufgrund dessen werden in der Regel zusätzliche Zielsequenzen durch neue TALENs anvisiert [194]. Aufgrund der bereits beschriebenen Probleme wurde jedoch auf die Herstellung weiterer TALENs verzichtet. Stattdessen wurde das zu diesem Zeitpunkt neu verfügbare CRISPR/Cas-System [143; 194] verwendet. Neben der deutlich einfacheren Herstellung ist das CRISPR/Cas-System auch nicht methylierungssensitiv [227] und überwindet eine weitere Limitierung der TALEN-Technologie.

### 4.1.2 CRISPR/Cas-vermittelte „Knock-Out“ Versuche

Aufgrund der vorangegangenen Problematik mit der TALEN-Technologie (vgl. 3.1.1) sollten zunächst Validierungsexperimente mit dem verwendeten CRISPR/Cas-System [53] durchgeführt werden. Hierzu wurde das Gen *ntl* aufgrund des gut sichtbaren Phänotypen verwendet, welcher sich durch das Fehlen der kaudalen Strukturen auszeichnet [200]. Die bei diesen Versuchen beobachteten hohen Mutationsraten von teilweise über 90 % in individuellen Embryonen zeigen deutlich die hohe Effizienz des CRISPR/Cas-Systems und stehen im Einklang mit bisherigen Publikationen [53; 148; 156; 228–232]. Aus diesen Ergebnissen lässt sich die Schlussfolgerung ziehen, dass das CRISPR/Cas-System erfolgreich etabliert und der hier angewandte Arbeitsablauf von der Herstellung der notwendigen RNAs, über die Mikroinjektion, bis zur Aufzucht und Genotypisierung der injizierten Embryonen erfolgreich angewendet werden konnte.

Nach Berichten über unspezifische Effekte des CRISPR/Cas-Systems [153; 154; 222; 231] wurden verschiedene Möglichkeiten zur Verhinderung solcher Effekte beschrieben. Einer dieser Möglichkeiten ist die Verwendung von „*truncated sgRNAs*“, die aufgrund ihrer geringeren Länge (<20 bp) eine höhere Spezifität aufweisen. Dies mag zunächst widersprüchlich erscheinen, da zunächst eine Verlängerung der sgRNAs logisch erscheint, um eine höhere Spezifität zu erreichen. Untersuchungen konnten jedoch zeigen, dass solche längeren sgRNAs auch eine verminderte Effektivität an der eigentlichen Zielsequenz aufweisen [146; 153; 157], während die Verwendung kürzerer sgRNAs oder solcher mit Fehlpaarungen im 5'-Bereich weiterhin zu effektiven Mutationen in der Zielsequenz führt [143; 154; 233]. Aus diesen Ergebnissen erwuchs die Hypothese, dass die Nukleotide des 5'-Bereich für die volle Aktivität der sgRNA nicht notwendig sind und normalerweise für Fehlpaarungen an andere Stelle innerhalb der Zielsequenz kompensieren. Somit sollte dann eine Verkürzung der sgRNA zu einer erhöhten Sensitivität gegenüber solchen Fehlpaarungen und somit auch zu einer erhöhten Spezifität führen. Tatsächlich konnte bei der Verwendung dieser „*truncated sgRNAs*“ eine deutliche Reduzierung der „*off-target*“-Effekte beobachtet werden [155]. Eine weitere Möglichkeit zur Verminderung unspezifischer Effekte ist die Verwendung von sogenannten CRISPR/Cas9-Nickasen. Damit ähnelt dieser Ansatz der TALEN-Technologie (vgl. Abbildung 1.7). Es kommen zwei unterschiedliche sgRNAs sowie Cas9-Nukleasen zum Einsatz, die lediglich einen Einzelstrangbruch einführen, so dass nur wenn beide sgRNAs in der korrekten Orientierung und Entfernung zueinander binden ein Doppelstrangbruch eingeführt werden kann [156]. Zwar erhöht diese Methode die Spezifität des CRISPR/Cas-Systems deutlich, jedoch erschwert sie zugleich die Suche nach geeigneten Zielregionen. Da aber für die sgRNA der in

dieser Arbeit verwendeten Zielsequenz (*stk33-l*) keine „*off-target*“-Effekte detektierbar waren, wurde dieses System hier nicht verwendet.

Die 48 Stunden nach erfolgter Injektion bei der Begutachtung der Embryonen beobachteten Unterschiede bezüglich der Überlebensrate, sowohl der injizierten, als auch der uninjizierten Embryonen, kann mit der unterschiedlichen Qualität der jeweiligen Gelege erklärt werden. Diese lässt sich aus dem Anteil der normal entwickelten Embryonen innerhalb der uninjizierten Kontrolle ablesen. Des Weiteren spielen die Beschaffenheit der Injektionsnadel und die damit erzeugten Verletzungen eine Rolle. Bei Gelegen, deren Überlebensrate bereits innerhalb der uninjizierten Kontrolle deutlich unter der 50 % lag und somit auch zu geringeren Überlebensraten bei den injizierten Embryonen führte, war diese hohe Mortalität auf hohe Nitrat- und Ammoniumwerte in den Aquarien, Krankheiten oder die wiederholte Verwendung der gleichen Weibchen für die Eiablage zurückzuführen [187]. Nach Behebung der jeweiligen Ursachen war innerhalb weniger Tage eine Normalisierung der Überlebensrate zu beobachten. Die deutlich niedrigeren Überlebensraten sowie die beobachteten Deformationen in injizierten Embryonen können zum einen auf die durch die Mikroinjektionsnadel verursachten Verletzungen zurückgeführt werden [234]. Zum anderen können unspezifische, dosisabhängige toxische Effekte der RNA ursächlich sein [234]. Dies wird durch die Beobachtung bestätigt, dass vergleichbare Überlebensraten und Deformationen auch nach alleiniger Injektion der Cas9 mRNA auftreten, welche mit einer Konzentration von 150 ng/μl im Gegensatz zur sgRNA mit nur 6,25 ng/μl den Großteil der injizierten RNA ausmacht (Daten nicht gezeigt). Diese Ergebnisse deuten auch darauf hin, dass die geringen Überlebensraten und die beobachteten Deformationen keinen spezifischen Effekt eines Verlustes von *stk33* darstellen. Hinzu kommt, dass im Gegensatz zu den vorangegangenen *ntl*-Versuchen (siehe 3.1.2) unter den beobachteten Deformationen kein spezifischer Phänotyp identifiziert werden konnte.

Bei der Genotypisierung der Embryonen konnten erfolgreich Mutationen innerhalb der Zielsequenz ohne detektierbare „*off-target*“-Effekte nachgewiesen werden. Bei der gelelektrophoretischen Analyse der Fragmente des T7E1-Assay sind jedoch nur das ursprüngliche PCR-Produkt (591 bp) sowie das größere der beiden Spaltfragmente (480 bp) sichtbar, während das kleinere Fragment (110 bp) vermutlich aufgrund der geringeren Fluoreszenz und abnehmenden Auflösung für kleinere Fragmente nicht mehr sichtbar ist. Um dieses dennoch nachweisen zu können, erfolgte eine Analyse der entstandenen Fragmente auf dem Bioanalyzer. Hier ist aufgrund der deutlich höheren Sensitivität auch das kleinere Spaltprodukt deutlich zu erkennen.

Der bei der gelelektrophoretischen Analyse des T7E1-Assay auftretende Schmier kann mit hoher Wahrscheinlichkeit auf eine unzureichende Spezifität der T7E1-Endonuklease

zurückgeführt werden, da schon bei einem leichten Überverdau ein vollständiger Abbau der DNA-Fragmente beobachtet werden kann (Daten nicht gezeigt). Das Auftreten der unspezifischen Produkte von ca. 210 bp und 370 bp kann so jedoch nicht erklärt werden. Da diese nicht zu erwartenden Produkte in allen analysierten Proben zu beobachten waren und zusammen in etwa der Größe des ursprünglichen PCR-Produktes von 591 bp entsprechen, handelt es sich hierbei aller Wahrscheinlichkeit nach um tatsächlich entstandene Spaltprodukte. Eine mögliche Erklärung hierfür ist das Vorkommen einer hypervariablen Region oder eines hypervariablen Nukleotids innerhalb des analysierten Bereiches. Bei der Analyse der beiden möglichen Schnittpunkte, die zu den beobachteten Spaltprodukten führen würden, zeigten sich im Bereich eines im Intron gelegenen Schnittpunktes tatsächlich zwei heterozygote SNPs (vgl. Abbildung A 7), während die Region um den zweiten möglichen Schnittpunkt keine Auffälligkeiten aufweist. Aufgrund dessen sind diese beiden SNPs wohl ursächlich für die beobachteten unerwarteten Spaltprodukte. Außerdem lässt sich so auch das Auftreten dieser Spaltprodukte in den uninjizierten Kontrollen erklären. Bei der gelelektrophoretischen Analyse der Produkte des T7E1-Assays zeigte sich weiterhin eine zusätzliche Bande oberhalb des ursprünglichen PCR-Produktes. Hierbei handelt es sich wahrscheinlich um nicht vollständig renaturierte, einzelsträngige DNA-Moleküle (ssDNA). Diese Vermutung wird dadurch bestätigt, dass diese bei der Analyse mittels Bioanalyzer nicht beobachtet werden können, da bei dem hier verwendeten Assay ein doppelstrangspezifischer DNA-Farbstoff zum Einsatz kommt (Auskunft Agilent Technologies).

Analog zu den *ntl*-Versuchen (3.1.2) bestätigen die durchgeführten Sanger-Sequenzierungen der *stk33-I* Zielregion ebenfalls die positiven Ergebnisse der T7E1-Assays. Auffällig ist hierbei jedoch, dass mit nur 16,5 % ein deutlich geringerer Anteil der Sequenzen Mutationen der *stk33-I* Zielregion enthielten (gegenüber 90 % bei den *ntl*-Versuchen). Diese geringere Ausbeute an mutanten Sequenzen kann auf zwei unterschiedliche Ursachen zurückgeführt werden. Zum einen weisen unterschiedliche sgRNA auch immer unterschiedliche Effizienzen auf [53; 229]. Zum anderen sind die deutlichen Unterschiede bezüglich der Ausbeute an mutanten Sequenzen zum größten Teil auf die phänotypische Selektion der Embryonen bei den *ntl*-Versuchen zurückzuführen. Da hier der Phänotyp bereits bekannt ist [200], war eine Selektion der Embryonen mit einem ausgeprägten *ntl*-Phänotyp für die Genotypisierung möglich. Bei den *stk33*-Versuchen war dies aufgrund des bisher noch nicht bekannten Phänotyps jedoch nicht möglich.

Im Anschluss an die erfolgreichen Injektionen wurden alle nach 48 Stunden normal entwickelten Embryonen bis zur Pubeszenz aufgezogen, um durch gezieltes Auskreuzen zunächst eine heterozygot mutante F<sub>1</sub>-Generation zu erhalten, aus welcher sich dann durch

Inzucht eine homozygot mutante (*stk33*<sup>-/-</sup>) F<sub>2</sub>-Generation erzeugen lässt (vgl. Abbildung 2.3). Jedoch erreichten nur ca. 9 % der injizierten Embryonen die Pubeszenz, während 80 % bis 90 % der uninjizierten Embryonen diese im Zuge der regelmäßigen Nachzuchten erreichten. Eine mögliche Erklärung hierfür ist, da es sich bei den erfolgreich injizierten Embryonen um genetische Mosaik handelt, dass nur Embryonen mit Mutationen in Geweben überlebt haben, in denen *stk33* keine Rolle während der Embryonalentwicklung spielt, oder dass eingeführten Mutationen nur heterozygot vorliegen. Da für *stk33* aufgrund früherer Untersuchungen bereits eine Rolle bei der Organogenese postuliert wurde [20], wäre es möglich, dass alle Embryonen mit *stk33* Mutationen in relevanten Organen und Geweben während der späteren Larvalstadien verstorben sind. Eine unmittelbare Genotypisierung der verstorbenen Larven war mit dem hier verwendeten Versuchsaufbau und aufgrund des schnellen Abbaus der verstorbenen Larven leider nicht möglich. Um diese Hypothese dennoch indirekt überprüfen zu können und um die überlebenden Fische genotypisieren zu können, wurden diese mit Wildtyp-Individuen verpaart und die hieraus hervorgegangenen Embryonen mittels T7E1-Assay und ggfs. durch eine Sanger Sequenzierung der Zielregion genotypisiert. In der Regel erfolgt die Genotypisierung der aus den injizierten Embryonen herangewachsenen F<sub>0</sub>-Generation direkt mit Hilfe des sogenannten „*fin clip*“ [192]. Hierbei wird den adulten Fischen ein Teil der Schwanzflosse amputiert und anschließend für die Genotypisierung verwendet, da diese vollständig regeneriert wird [187]. Im Rahmen dieser Arbeit wurde aber aufgrund der tierschutzrechtlichen Bestimmungen sowie der geringeren Aussagekraft des „*fin clip*“ Abstand genommen. Da die adulten Individuen der F<sub>0</sub>-Generation in der Regel genetische Mosaik darstellen, erlaubt eine möglicherweise mutante Schwanzflosse nur selten eine Aussage über die genetische Beschaffenheit der für eine Vererbung der eingeführten Mutationen notwendigen Gonaden. Ohne das entsprechende Tier für eine direkte Genotypisierung der Gonaden töten zu müssen, ist dies nur indirekt durch die Genotypisierung der aus der Auskreuzung entstandenen F<sub>1</sub>-Generation möglich. Interessanterweise wies jedoch keine der aus diesen Auskreuzungen entstandenen F<sub>1</sub>-Generationen Mutationen innerhalb der *stk33*-I Zielregion auf. Hieraus lässt sich schließen, dass trotz der erfolgreichen Genotypisierungen 48 Stunden nach erfolgter Injektion keiner der überlebenden Zebrafische der F<sub>0</sub>-Generation über mutante Gonaden verfügte. Um zu überprüfen, ob diese Fische möglicherweise in anderen Organen oder Geweben Mutationen tragen, oder ob die bis zur Pubeszenz herangewachsenen Fische über keinerlei Mutationen verfügen, wurden diese anschließend vollständig genotypisiert. Die Tatsache, dass sich auch in den übrigen Geweben keinerlei Mutationen nachweisen ließen, bekräftigt die zuvor aufgestellte Hypothese, dass alle mutanten Fische in den späteren Larvalstadien verstorben sind und nur nicht mutante Fische

bis zur Pubeszenz heranwachsen konnten. Normalerweise wäre hier jedoch zu erwarten gewesen, dass zu mindestens Tiere mit nur heterozygoten Mutationen überlebensfähig sein sollten. Dies würde eine bereits vermutete Rolle von *stk33* bei der Organogenese [20] deutlich bekräftigen. Eine mögliche Erklärung für das Fehlen solcher heterozygoten Mutationen könnte sein, dass durch die hohe Effizienz des CRISPR/Cas-Systems ein Großteil der Mutation homozygot vorliegt. Weiterhin wäre ein Gendosis-Effekt denkbar, welcher bereits bei einer heterozygoten Mutation von *stk33* zu einem Funktionsverlust der STK33-basierten Prozesse führen könnte. Ohne weitere Untersuchungen diesbezüglich bleiben diese Vermutungen jedoch spekulativ.

Um dies überprüfen zu können, wären umfangreichere Injektionen und eine direkte Genotypisierung der verstorbenen Larven notwendig. Aufgrund der vergleichsweise langen Generationszeit des Zebrafisches von drei bis vier Monaten war dies im Rahmen dieser Arbeit jedoch leider nicht mehr möglich. Eine Möglichkeit um ein Sterben der mutanten Larven zu verhindern, könnte ein gewebsspezifisches Ausschalten von *stk33* in den Gonaden sein. Solche gewebsspezifischen CRISPR/Cas-Systeme wurden bereits beschrieben und etabliert [235]. Auf diese Weise könnten Fische erzeugt werden, die Mutationen für *stk33* ausschließlich innerhalb der Gonaden aufweisen. Da alle anderen Organe und Gewebe nicht von diesen Mutationen betroffen wären, sollten diese Tiere in der Lage sein, ohne Probleme bis zur Pubeszenz heranzuwachsen und im Falle einer Auskreuzung heterozygot mutanten Nachwuchs zu erzeugen.



## 4.2 *Stk33* im Reich *Animalia*

Im Anschluss an die erfolgreiche Identifizierung von *stk33* in den Urkiefervögeln (3.3.1) bzw. dessen unitären Pseudogens in Neukiefervögeln (3.3.2 & 3.3.3), sollte im Folgenden untersucht werden, ob *stk33* auch außerhalb der Wirbeltiere in weiteren Vertretern des Reiches *Animalia* vorhanden ist und ob es ggfs. im Laufe der Evolution in weiteren Taxa verloren gegangen ist. Aufgrund der Vielzahl zu analysierender Arten, kam hierbei ein reiner „*data-mining*“ Ansatz auf genomische sowie auf transkriptomische Daten in öffentlich zugänglichen Datenbanken (vgl. 2.5.6) zum Einsatz. „*Data-mining*“ oder auch „*knowledge discovery in databases*“ (KDD) beschreibt die systematische Auswertung bereits vorhandener Daten mit dem Ziel, daraus neue Erkenntnisse zu gewinnen [236]. Dies stellt aufgrund der exponentiell zunehmenden Verfügbarkeit von Daten [237] eine attraktive Alternative zu klassischen labortechnischen Untersuchungen dar. So enthält allein das „*Sequence Read Archive*“ des NCBI (SRA; <http://www.ncbi.nlm.nih.gov/Traces/sra>) Sequenzinformationen von über 4,6 Petabasen ( $4,6 \times 10^{15}$  Basen; Stand 01/2016). Mit Hilfe des unter 2.5.6 beschriebenen Ansatzes konnten so insgesamt 726 Datensätze von 488 verschiedenen Tierarten analysiert werden. Bei 189 der analysierten Datensätze handelt es sich um genomische Datensätze und bei den verbleibenden 537 um transkriptomische Datensätze.

Um das Vorkommen der Serin/Threonin Kinase 33 im Reich *Animalia* zu untersuchen, wurden zunächst verschiedene Vertreter der den Wirbeltieren (*Vertebrata*) am nächsten verwandten Chordatiere untersucht. Diese umfassen neben den Wirbeltieren auch die Manteltiere (*Tunicata* oder *Urochordata*) sowie die zu den Schädellosen (*Cephalochordata*) gehörenden Lanzettfischchen. Hierbei ist es interessant, dass nicht die Schädellosen, sondern die Manteltiere trotz ihrer stark divergenten Morphologie und Lebensweise das den Wirbeltieren am nächsten verwandte Taxon darstellen [202]. Weiterhin wurden innerhalb der Neumünder (*Deuterostomia*) die Kiemenlochtiere (*Hemichordata*) und die Stachelhäuter (*Echinodermata*) analysiert. Die Ergebnisse dieser Untersuchungen konnten eindeutig das Vorhandensein und die Expression von *stk33* in diesen Stämmen nachweisen (vgl. Tabelle 3.2 und Abbildung 4.1). Die Analyse der aus den genomischen bzw. transkriptomischen Contigs abgeleiteten Proteinsequenzen mit Hilfe von Proteinalignments und InterPro zeigen zudem eindeutig, dass mit dem Vorhandensein der Proteinkinasedomäne, der ATP-Bindestelle sowie dem aktiven Zentrum der Serin/Threonin spezifischen Kinasen (vgl. Abbildung 3.5 & Abbildung 3.6) alle Voraussetzungen für eine funktionale Kinase gegeben sind.

Aufgrund der starken Syntanie der gesamten *stk33*-flankierenden Region in Vertebraten, welche sich zwischen den Genen *st5* und *ric3* erstreckt, sollte auch in den nahen verwandten

Manteltieren und den zu den Schädellosen gehörenden Lanzettfischchen untersucht werden, ob die genomische Region um *stk33* herum hier ebenfalls diese Syntänie aufweist. Hierbei zeigte sich, dass in den Manteltieren und Lanzettfischchen diese Syntänie nicht mehr besteht und *stk33* in diesen Arten von gänzlich anderen Genen flankiert wird (vgl. Abbildung 3.4). Genomweite BLASTx-Suchen konnten zwar einige der im Menschen *stk33* flankierenden Gene in diesen Organismen identifizieren, jedoch befinden diese sich nicht benachbart innerhalb einer Chromosomenregion, sondern sind über das gesamte Genom verteilt. Für die Tatsache, dass nicht alle flankierenden Gene in Manteltieren und Schädellosen identifiziert werden konnten, gibt es verschiedene Erklärungen. Zum einen kann ein Teil dieser nicht gefundenen Gene erst an der Basis der Wirbeltiere neu entstanden oder in den Manteltieren und Schädellosen verloren gegangen sein. Ein Verlust dieser Gene erscheint vor allem für die Vertreter der Manteltiere plausibel, da diese im Verlauf ihrer Evolution eine erhebliche Reduzierung der Genomgröße und somit auch den Verlust zahlreicher Gene erfahren haben [238–241]. Zum anderen kann nahezu keines der publizierten Genome als tatsächlich vollständig angesehen werden, da es in der Regel aus vielen tausend verschiedenen Contigs bzw. Scaffolds besteht, wobei die genetische Information aus den Lücken zwischen den einzelnen Contigs nicht zugänglich ist. Der Grund hierfür liegt in der Assemblierung von z.B. repetitiven Bereichen mit Hilfe von vergleichsweise kurzen Sequenzen aus den verschiedenen Hochdurchsatz-Sequenzieretechnologien [242; 243]. Der Anteil dieser nicht zugänglichen Information schwankt stark, kann aber teilweise mehr als 10 % mit Bezug auf das gesamte Genom betragen [244]. Hier wäre es möglich, dass einige der *stk33* flankierenden Gene in solchen Lücken liegen und somit in den aktuellen Versionen der jeweiligen Genome nicht auffindbar sind, obwohl diese eigentlich vorhanden sind. Um dies ausschließen zu können, wäre zunächst eine Zuordnung der Contigs zu den jeweiligen Chromosomen und anschließend die Verknüpfung der zu einem Chromosom gehörenden Contigs notwendig. Dies kann bis zu einem gewissen Punkt bioinformatisch über die Verwendung spezieller Programme [245] erfolgen. Die nach diesem Schritt noch verbleibenden Lücken können dann über gezielte Sanger-Sequenzierungen (ggfs. BAC-Sequenzierung) [244] oder mit Hilfe der vergleichsweise langen Sequenzen der neuen „*single molecule real time*“ (SMRT) Sequenzieretechnologie von Pacific Biosciences [246; 247] geschlossen werden. Erst mit Hilfe von weitgehend vollständigen Genomen lässt sich dann eine fundierte Aussage darüber treffen, ob die nicht gefundenen Gene des *stk33* Locus tatsächlich in den Manteltieren und Schädellosen fehlen oder aufgrund der unvollständigen Genomsequenzen nicht auffindbar sind. In vielen anderen Chordaten war die Analyse des *stk33* umgebenden Locus nicht möglich, da hier *stk33* zwar identifiziert werden konnte, der *stk33* enthaltende Contig aber i.d.R. zu kurz war, um eine fundierte Aussage über

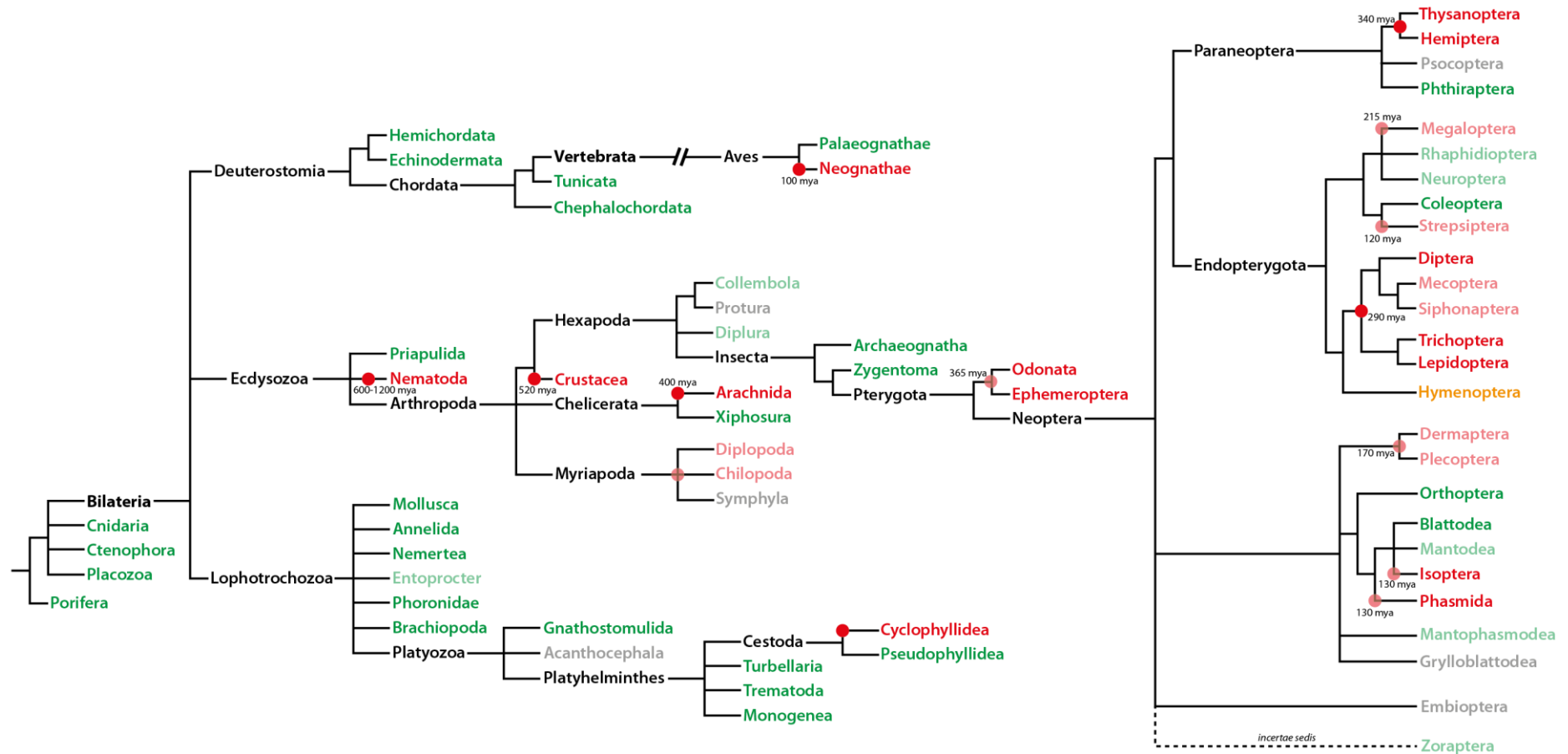
die benachbarten Gene treffen zu können. Hieraus lässt sich schließen, dass *stk33* in dem gemeinsamen Vorfahren der Wirbeltiere nicht in eine schon existierende Chromosomenregion mit den Genen *ric3*, *lmo1*, *trim66*, *rpl27a* und *st5* transloziert wurde. Stattdessen ist davon auszugehen, dass der noch heute in allen Wirbeltieren zu findende syntäne Locus um *stk33* herum (vgl. Abbildung 3.10 A) mit hoher Wahrscheinlichkeit durch tiefgreifende genomische Rearrangements in dem gemeinsamen Vorfahr aller Wirbeltiere entstanden sein muss. Tatsächlich konnten durch vergleichende Genomanalysen zwischen Wirbeltieren und Lanzettfischchen solche massiven genomischen Rearrangements in den frühen Wirbeltieren nachgewiesen werden [248].

Die weiteren Analysen innerhalb des Reiches *Animalia* konnten *stk33* auch außerhalb der *Chordata* in zahlreichen Taxa der *Deuterostomia* (3.2.1), *Lophotrochozoa* (3.2.2), *Ecdysozoa* (3.2.3) sowie in allen basalen Tierstämmen (3.2.4) sowohl auf genomischer, als auch auf transkriptomischer Ebene nachweisen. Zwar konnten hierbei in einigen Fällen aufgrund von kurzen Contigs oder Lücken innerhalb der Assemblierungen nur partielle *stk33* Sequenzinformationen gewonnen werden. Jedoch enthalten die identifizierten kodierenden Exons keinerlei Indels oder Stopkodons innerhalb des offenen Leserahmens, so dass davon ausgegangen werden kann, dass der jeweilige Organismus über ein funktionsfähiges STK33 verfügt. Ansonsten wären bei einer Inaktivierung von *stk33*, analog zu den Untersuchungen der Neukiefervögel (vgl. 3.3.3), solche Mutationen in den identifizierten Exons zu erwarten gewesen. Hierbei ist es besonders interessant, dass *stk33* von komplexen Säugetieren bis hin zu den strukturell sehr einfachen, basalen Tieren wie den Nesseltieren (*Cnidaria*), Rippenquallen (*Ctenophora*), Schwämmen (*Porifera*) und den strukturell einfachsten aller vielzelligen Tiere, den Plattentieren (*Placozoa*) [211], nachgewiesen werden konnte. Diese weite Verbreitung spricht für eine grundlegende Funktion von *stk33* innerhalb der vielzelligen Tiere (*Metazoa*).

Auch hier konnten mit dem Nachweis der Proteinkinasedomäne, der ATP-Bindestelle sowie dem aktiven Zentrum der Serin/Threonin spezifischen Kinasen alle notwendigen Bestandteile für eine funktionale Kinase durch die Analyse der Aminosäuresequenzen mit Hilfe eines multiplen Sequenzalignments (Abbildung 3.6) und InterPro (Abbildung 3.5) identifiziert werden. Zwar zeigt das Alignment der verschiedenen STK33 Proteine innerhalb der ATP-Bindestelle nur eine moderate Konservierung, jedoch fällt auch auf, dass einige der Glycine am N-terminalen Ende sowie das Lysin am C-terminalen Ende der ATP Bindestelle in allen Proteinsequenzen konserviert vorliegen (siehe Abbildung 3.6). Bei diesen Aminosäuren, insbesondere bei dem Lysin, handelt es sich um die für die Bindung von ATP verantwortlichen Aminosäuren (Expasy prosite Eintrag PS00107). Daher kann davon ausgegangen werden, dass

trotz der nur moderaten Konservierung über den gesamten Bereich in allen identifizierten Proteinen eine funktionale ATP-Bindestelle vorhanden ist. Hierfür spricht auch, dass in jeder dieser Proteinsequenzen die ATP-Bindestelle bei der Analyse mittels InterPro erkannt wird. Bereits in dem Alignment der in den Urkiefervögeln gefundenen STK33 Proteine deutet sich an, dass die von Mujica *et al.* beschriebene Asx-reiche Schleife nicht konserviert ist. Noch deutlicher wird dies durch das Alignment der im Reich *Animalia* identifizierten STK33-Proteine. Hier ist keinerlei Konservierung im Vergleich zu den STK33 Proteinsequenzen von Mensch und Maus mehr zu erkennen (siehe Abbildung 3.6). Diese Asx-reiche Schleife wurde ursprünglich von Mujica in unmittelbarer Nachbarschaft zum aktiven Zentrum identifiziert und als hoch konservierter und charakteristischer Abschnitt von polaren und sauren Aminosäuren, spezifisch für STK33, mit unbekannter Funktion beschrieben [19]. Die hier gewonnenen Erkenntnisse legen jedoch nahe, dass diese Asx-reiche Schleife kein spezifisches Merkmal aller STK33-Proteine darstellt. Dies macht auch eine mögliche Funktionalität unwahrscheinlich.

Trotz des erfolgreichen Nachweises von *stk33* innerhalb der *Metazoa* zeigte sich auch, dass neben den Neukiefervögeln *stk33* in zahlreichen weiteren Taxa trotz der teilweise umfassenden Verfügbarkeit von genomischen und transkriptomischen Daten nicht nachgewiesen werden konnte (vgl. Abbildung 4.1). Im Unterschied zu den Untersuchungen der Neukiefervogel (vgl. 3.3.3) konnten hier jedoch keinerlei Überreste der Serin/Threonin Kinase 33 identifiziert werden. Dies ist höchst wahrscheinlich mit der längeren Zeitspanne seit dem Verlust von *stk33* im Vergleich zu der Inaktivierung von *stk33* in Neukiefervögeln vor ca. 100 Millionen Jahren zu erklären. Bei einer ausreichend langen Zeit ohne Selektionsdruck können sich in einem Locus derart viele Mutationen akkumulieren, dass etwaige Überreste von *stk33* über die Suche nach Homologien (BLAST) nicht mehr identifiziert werden können. Dies ist vor allem für das vergleichsweise gut abgesicherte Fehlen von *stk33* in den Fadenwürmern (*Nematoda*) vor 600 bis 1200 Millionen Jahren [249–252], den Krebstieren (*Crustacea*) vor ca. 520 Millionen Jahren [253] und den Spinnentieren (*Arachnida*) vor ungefähr 400 Millionen Jahren [254] plausibel. Auch in einigen Insektenordnungen lässt sich das vollständige Fehlen von *stk33* Überresten durch die lange Zeitspanne seit der Inaktivierung erklären (vgl. Abbildung 4.1) [255]. Jedoch finden sich auch Ordnungen wie beispielsweise die Gespenstschrecken (*Phasmida*) und die Termiten (*Isoptera*), in denen *stk33* vermutlich erst vor 120 bis 170 Millionen Jahren verloren gegangen ist [255]. Hier ist es fraglich, ob diese Zeitspanne ausreichend für solch starke Veränderungen innerhalb des *stk33* Locus ist. Hier sind erhöhte Mutationsraten oder ggfs. die vollständige Deletion des *stk33* Locus plausible Erklärungen.



**Abbildung 4.1 | Kladogramm aller untersuchten Taxa aus dem Reich *Animalia*.** Taxa mit einem positiven Befund bezüglich des Vorhandenseins von *stk33* sind in grün, Taxa mit einem negativen Befund sind in rot dargestellt. Wenn aufgrund der schlechten Datenlage keine zuverlässige Aussage möglich war, sind die betroffenen Taxa ausgeblast. Taxa mit nur ein oder zwei verfügbaren transkriptomischen Datensätzen sind grau markiert, da diese im Falle eines negativen Befundes keine Aussage über das Vorhandensein von *stk33* zulassen. Die potenziellen Verluste von *stk33* sind innerhalb des Kladogramm durch rote Punkte gekennzeichnet, die Zeitpunkte der Inaktivierung sind in Millionen Jahren (mya) angegeben. Die Vertreter der Wirbeltiere sind, mit Ausnahme der Klasse *Aves*, in diesem Kladogramm nicht dargestellt, da hier das Vorhandensein von *stk33* bereits nachgewiesen wurde und nicht Bestandteil dieser Arbeit waren.

Eine Besonderheit innerhalb der hier durchgeführten Analysen stellt die Ordnung der Hautflügler (*Hymenoptera*) dar. Hier wurden insgesamt 54 verschiedene Arten anhand von 30 genomischen sowie 52 transkriptomischen Datensätzen analysiert. In lediglich zwei der 54 analysierten Arten, *Fopius arisanus* und *Athalia rosae*, konnte ein Nachweis für *stk33* geführt werden. Dieser erfolgte sowohl auf genomischer als auch auf transkriptomischer Ebene, während in allen anderen analysierten Datensätzen keinerlei Hinweise auf *stk33* zu finden waren. Dies ist besonders bei der Betrachtung der deutlich unterschiedlichen Position dieser beiden Arten innerhalb des Stammbaumes der Hautflügler interessant (vgl. Abbildung A 11). Während *Fopius arisanus* zu der Familie der Brackwespen (*Braconidae*) innerhalb der Unterordnung der Taillenwespen (*Apocrita*) gehört, ist *Athalia rosae* ein Mitglied der Familie der Echten Blattwespen (*Tenthredinidae*) innerhalb der Unterordnung der Pflanzenwespen (*Symphyla*) [256]. Innerhalb der Familie der Brackwespen wurden neben *Fopius arisanus* noch weitere Arten sowohl auf genomischer als auch auf transkriptomischer Ebene ohne einen Hinweis auf *stk33* untersucht. Hieraus lässt sich schließen, dass *stk33* wahrscheinlich nur innerhalb der Gattung *Fopius* erhalten geblieben ist. Für die Familie der Echten Blattwespen konnte aufgrund fehlender Daten nur eine weitere Art (*Tenthredo koehleri*, vgl. Tabelle A 1) auf transkriptomischer Ebene ohne Hinweise auf *stk33* analysiert werden. Aufgrund dessen ist es hier nicht klar, ob *stk33* nur innerhalb der Gattung *Fopius* oder innerhalb der kompletten Familie der Echten Blattwespen erhalten geblieben ist. Um hierüber eine genauere Aussage treffen zu können sind weitere Untersuchungen mit zusätzlichen genomischen und transkriptomischen Daten notwendig.

Ein generelles Problem solcher Analysen ist, dass bei dem vollständigen Fehlen von identifizierbaren *stk33* Überresten eine fundierte Aussage über einen eventuellen Verlust von *stk33* innerhalb eines Taxon erst mit einer guten Sequenzabdeckung des jeweiligen Taxons möglich wird. Finden sich hingegen mit Indels und Stopkodons angereicherte Überreste, vergleichbar zu den Ergebnissen der Analyse der Unterklasse der Neukiefervögel (siehe 4.3.2 und insbesondere Abbildung 3.11), ist die Degeneration und damit der Funktionsverlust eines Genes unmittelbar nachweisbar. Nicht desto trotz konnte beispielweise innerhalb der Fadenwürmer, der Krusten- und Spinnentiere, der Zweiflügler (*Diptera*) sowie weiterer Taxa aufgrund der ausreichenden Verfügbarkeit von genomischen und transkriptomischen Daten der Verlust von *stk33* sicher nachgewiesen werden. Doch insbesondere in der Klasse der Insekten finden sich zwar innerhalb mehrerer Ordnungen Hinweise auf einen möglichen Verlust von *stk33*, jedoch kann dieser aufgrund der unzureichenden Verfügbarkeit von Daten nicht zweifelsfrei nachgewiesen werden (vgl. Abbildung 4.1). Dies ist insbesondere der Fall, wenn ausschließlich transkriptomische Daten für die Analyse einer Ordnung zur Verfügung

stehen. Für nicht ubiquitär exprimierte Gene, wie beispielsweise *stk33*, kann in Abhängigkeit von Gewebe oder Stadium, welches dem jeweiligen Datensatz zu Grunde liegt, ggfs. gar keine Information bezüglich dessen Expression erwartet werden. Auch enthielten einige der bereits assemblierten Transkriptome nur einige hundert oder wenige tausend Sequenzen, sodass diese Datensätze keine vollständige Repräsentation des Transkriptoms darstellen. Aus diesen Gründen konnte beispielsweise *stk33* in allen untersuchten Genomen, aber nur in 12 der 25 analysierten transkriptomischen Datensätze innerhalb der Ordnung der Käfer (*Coleoptera*) nachgewiesen werden (Tabelle 3.4). Aus rein negativen Befunden der transkriptomischen Analysen lässt sich also noch kein Verlust von *stk33* ableiten. Hierfür sind zusätzlich genomische Informationen notwendig, die jedoch für viele der untersuchten Ordnungen trotz der exponentiell anwachsenden Datenmenge [237] noch nicht vorliegen. Um bei analysierten Ordnungen mit nicht eindeutig positiven Befunden und geringer, ausschließlich transkriptomischer Abdeckung (vgl. Abbildung 4.1 und Tabelle 3.4) dennoch eine fundierte Aussage treffen zu können, wäre die Überprüfung der ggfs. identifizierten Contigs mit Hilfe von PCR-Analysen notwendig. Dies stellt bei der Vielzahl der zu analysierenden Arten jedoch einen erheblichen labortechnischen Aufwand dar und insbesondere die Beschaffung geeigneten Gewebematerials kann je nach Art problematisch sein.

Trotz all dieser Probleme und den daraus resultierenden Limitierungen eines solchen reinen „*data-mining*“ Ansatzes lässt sich die Schlussfolgerung ziehen, dass *stk33* neben den Wirbeltieren auch über das gesamte Reich *Animalia* verteilt vorkommt und somit in dem gemeinsamen Vorfahren aller vielzelligen Tiere (*Metazoa*) vor ca. 1,5 Milliarden Jahren [257] vorhanden gewesen sein muss. Des Weiteren muss *stk33* mehrmals im Laufe der Evolutionsgeschichte der vielzelligen Tiere verloren gegangen sein. Anhand der in dieser Arbeit durchgeführten Analysen lassen sich insgesamt mindestens sieben voneinander unabhängige Verluste von *stk33* nachweisen (vgl. Abbildung 4.1). Die Datenlage deutet auf noch weitere Verluste hin, jedoch sind hier aufgrund der unzureichenden Verfügbarkeit geeigneter Datensätze für eine fundierte Aussage weitere Analysen notwendig.

### 4.3 *Stk33* in der Klasse *Aves*

Ein Nachweis von *stk33* erfolgte bisher ausschließlich innerhalb des Unterstammes der Wirbeltiere (*Vertebrata*). So konnte *stk33* u.a. in den großen Klassen der Säugetiere (*Mammalia*), Reptilien (*Reptilia*), Amphibien (*Amphibia*) sowie in den Knorpel (*Chondrichthyes*) und Knochenfischen (*Osteichthyes*) nachgewiesen werden (vgl. Abbildung A 10). Lediglich innerhalb der Klasse der Vögel (*Aves*) konnte solch ein Nachweis bisher nicht geführt werden. Diese ist unterteilt in die Unterklassen der Urkiefervögel (*Palaeognathae*), welche die flugunfähigen Laufvögel und die Steiße hühner umfasst und mit nur 58 bekannten Arten [216] lediglich 0,5 % aller rezenten Vogelarten beinhaltet, sowie die Unterklasse der Neukiefervögel (*Neognathae*), die alle übrigen rezenten Arten beinhaltet [215–217]. Sie umfasst ca. 10.500 rezente Arten [258], wobei ein Großteil dieser Diversität durch eine rasche Ausbreitung der Arten als Folge eines Massenaussterbens vor ca. 66 Millionen Jahren während des Übergangs von der Kreidezeit zum Paläogen entstand [215; 259].

Aufgrund des fehlenden Nachweises von *stk33* innerhalb der Klasse *Aves* war es ein Ziel dieser Doktorarbeit, diese auf das Vorhandensein von *stk33* zu untersuchen. Hierzu wurde zunächst das Haushuhn (*Gallus gallus domesticus*), als Modellorganismus innerhalb der Klasse *Aves*, im Rahmen der Diplomarbeit von Daniel Schäfer [212] untersucht. Hierbei konnten keinerlei Hinweise für das Vorhandensein eines funktionsfähigen Gens für *STK33* gefunden werden. Da *stk33* in vielen anderen Wirbeltier-Klassen bereits nachgewiesen werden konnte, stellte sich daraufhin die Frage, ob *stk33* in Vögeln generell nicht vorhanden ist oder das Haushuhn hier eine Ausnahme darstellt. Um diese Frage beantworten zu können, wurde zunächst im Rahmen der Diplomarbeit von Steffen Lerch [213] die Analyse des Afrikanischen Vogelstrauß (*Struthio camelus*), dem basalsten rezenten Vertreter der Klasse *Aves*, begonnen und im Rahmen dieser Doktorarbeit vervollständigt (vgl. 3.3.1). Im weiteren Verlauf dieser Doktorarbeit wurden die Transkriptomte und Genome mittels einer Kombination von „Next-Generation-Sequencing“ Daten und PCR von über 30 verschiedenen Vogelarten auf die Präsenz bzw. Expression von *stk33* hin untersucht. Die im Rahmen der vorliegenden Arbeit erzielten Ergebnisse bezüglich des Vorhandenseins von *stk33* in der Klasse *Aves* konnten bereits erfolgreich in einer Fachzeitschrift mit Peer-Review Gutachten publiziert werden [260].



### 4.3.1 *Stk33* in Urkiefervögeln (*Palaeognathae*)

Die Untersuchung der Unterklasse der Urkiefervögel auf ein eventuelles Vorhandensein von *stk33* erfolgte für den Afrikanischen Vogelstrauß (*Struthio camelus*) auf genomischer und transkriptomischer Ebene, für den Weißkehltinamu (*Tinamus guttatus*) auf genomischer und für den Emu (*Dromaius novaehollandiae*) auf transkriptomischer Ebene. Aufgrund der limitierten Verfügbarkeit von Daten konnten insgesamt nur drei Vertreter der Unterklasse *Paleognathae* analysiert werden. Da diese drei Arten aber die beiden Ordnungen der Laufvögel (*Struthioniformes*) und Steiþhühner (*Tinamiformes*) abdecken [216] kann die Analyse dieser drei Spezies als repräsentativ für alle Urkiefervögel angesehen werden.

Die Ergebnisse dieser Untersuchungen (siehe 3.3.1) konnten deutlich das Vorhandensein eines funktionalen *stk33*-Gens auf genomische Ebene, sowohl im Afrikanischen Vogelstrauß als auch im Weißkehltinamu zeigen. Hierbei konnte für den Vogelstrauß ein Contig mit insgesamt 14 Exons identifiziert werden, während für den Weißkehltinamu lediglich 13 Exons nachgewiesen werden konnten. Hier fehlt das Exon 13, während die flankierenden Exons 12 und 14 jedoch vorhanden sind. Eine mögliche Erklärung hierfür könnten Fehler innerhalb der genomischen Assemblierung sein, allerdings finden sich zwischen den flankierenden Exons 12 und 14 keine Lücken innerhalb der Assemblierung und auch eine Kartierung der genomischen Daten gegen diesen Bereich zeigten keine Hinweise auf mögliche Fehler in der Assemblierung. Zwar liegt das im Weißkehltinamu fehlende Exon 13 im 3'-Bereich des offenen Leserahmens, doch liegt es innerhalb des schlecht konservierten Bereiches abseits der Kinasedomäne und keines der identifizierten Exons trägt Leserasterverschiebende Indels oder frühzeitige Stopkodons innerhalb der kodierenden Region (vgl. Abbildung 3.11). Daraus lässt sich schließen, dass auch der Weißkehltinamu über ein funktionsfähiges *stk33* verfügt.

Des Weiteren konnte die Expression des identifizierten *stk33* für den Vogelstrauß anhand von RT-PCRs sowie der Analyse von transkriptomischen NGS-Daten zweifelsfrei nachgewiesen werden. Für den Weißkehltinamu war dies aufgrund der mangelnden Verfügbarkeit von transkriptomischen Daten nicht möglich. Die Analyse des Emus erfolgte, aufgrund des Fehlens geeigneter genomischer Datensätze, ausschließlich auf transkriptomischer Ebene. Hierbei konnte die Existenz einer *stk33* mRNA zweifelsfrei nachgewiesen werden.

Die Analyse der aus den identifizierten genomischen Contigs bzw. mRNA Sequenzen abgeleiteten Proteinsequenzen mittels BLASTx belegt eindeutig, dass es sich bei diesen Proteinen um STK33 handelt. Außerdem zeigen die funktionellen Analysen mittels InterPro klar, dass mit der Serin/Threonin spezifischen Kinasedomäne, der ATP-Bindestelle und dem aktiven Zentrum der Serin/Threonin-spezifischen Kinasen (vgl. Abbildung 3.8) alle notwendigen

Bestandteile einer funktionalen Proteinkinase vorhanden sind. Dies wird durch das Alignment der in der Klasse *Aves* gefundenen Proteine mit den STK33 Proteinsequenzen von Mensch, Maus, Rotkehlantilope und Zebrafisch, insbesondere durch den hohen Grad der Konservierung innerhalb des aktiven Zentrums, bestätigt (siehe Abbildung 3.9).

Auffällig ist hier auch, dass die STK33-Proteine der Urkiefervögel in dem Bereich der von Mujica beschriebenen Asx-reichen Schleife [19] nicht mehr konserviert zu sein scheinen. Handelt es sich in Mensch und Maus noch bei sechs der neun Aminosäuren um Asparagin oder Asparaginsäure, so sind es in den STK33-Proteinsequenzen der Urkiefervögel gerade noch drei der neun Aminosäuren (Abbildung 3.9). Dies legt die Vermutung nahe, dass es sich hierbei nicht um eine spezifische Besonderheit von STK33 handelt. Zusammenfassend lässt sich aus diesen Ergebnissen die Schlussfolgerung ziehen, dass ein funktionales und exprimiertes *stk33* in allen Urkiefervögeln vorhanden ist.

#### 4.3.2 *Stk33* in Neukiefervögeln (*Neognathae*)

Innerhalb der Neukiefervögel wurde zunächst das Haushuhn (*Gallus gallus*) auf ein mögliches Vorhandensein von *stk33* hin untersucht. Trotz des Vorhandenseins der in den Wirbeltieren konservierten Chromosomenregion um *stk33*, konnten lediglich nicht funktionale Überreste von *stk33* identifiziert werden. Eine mögliche Erklärung hierfür könnten Fehler innerhalb der Genomsequenz sein. Aufgrund der hohen Qualität des Genoms des Huhns sowie der genomischen Sequenz kann eine solch fehlerhafte Genomsequenz jedoch ausgeschlossen werden [212]. Da auch die transkriptomischen Analysen und auch eine genomweite Suche keine Hinweise auf *stk33* lieferten, führten diese Ergebnisse zu der Schlussfolgerung, dass *stk33* im Haushuhn nur noch als funktionsloses, unitäres Pseudogen vorliegt.

Um zu überprüfen, ob das Haushuhn hier eine Ausnahme darstellt und andere Neukiefervögel möglicherweise über eine intakte Kopie von *stk33* verfügen, oder ob auch hier *stk33* nur noch als unitäres Pseudogen vorliegt, wurden die Genome und Transkriptome von insgesamt 27 weiteren Neukiefervögeln untersucht. Jedoch konnten auch hier lediglich nichtfunktionale Überreste von *stk33* gefunden werden. Innerhalb des genomischen *stk33*-Locus zeigten sich ebenfalls mehrere große genomische Deletionen über mehrere Kilobasen, die zu dem Verlust zahlreicher Exons führten. Hierbei konnten auch systematische Übereinstimmung zwischen verwandten Arten identifiziert werden. So konnte in allen untersuchten Arten der Sperlingsvögel (*Passeriformes*) der Verlust der Exons 5, 8, 12 und 13 gezeigt werden (vgl. Abbildung A 9). Außerdem fanden sich in den verbleibenden Exons auch zahlreiche Indels sowie vorzeitige Stopkodons (vgl. Abbildung 3.11 & Abbildung A 9). Diese Mutationen müssen

zweifelsohne zum Funktionsverlust der *stk33* Überreste in Neukiefervögeln geführt haben. Die durchgeführten transkriptomischen Untersuchungen konnten des Weiteren keinerlei *stk33*-relevante Transkripte nachweisen. Die durchgeführten BLASTn/x-Suchen der Genome aller untersuchten Neukiefervögel zeigen deutlich, dass hier keine Genduplikation stattgefunden hat. Auch konnte ausgeschlossen werden, dass *stk33* neue Exons erworben hat („*exon gain*“). Eine weitere Möglichkeit ist, dass ein *stk33* Transkript zurück in das Genom retrotransponiert wurde und so das originale *stk33* funktionell substituierte. Jedoch werden diese als prozessierte Pseudogene bezeichneten genetischen Elemente aufgrund des nicht zielgerichteten Einbaus in das Genom und des damit einhergehenden Fehlens eines geeigneten Promotors als nicht funktional angesehen („*dead on arrival*“) [261; 262]. BLASTn/x-Suchen mit allen bekannten *stk33*-mRNA und Proteinsequenzen zeigten weiterhin keinerlei Hinweise auf solch eine Retrotransposition. Diese Ergebnisse führen zu der Schlussfolgerung, dass *stk33* nicht nur im Haushuhn, sondern auch in allen anderen Neukiefervögeln im Verlauf der Evolutionsgeschichte der Klasse *Aves* zu einem unitären Pseudogen mutiert ist. Somit stellt die Unterklasse der Neukiefervögel die einzige Gruppe innerhalb der Wirbeltiere ohne ein funktionsfähiges *stk33* dar (Abbildung A 10). Aufgrund stetig sinkender Sequenzierungskosten, werden immer mehr groß angelegte Sequenzierungsprojekte initiiert. Hierzu zählt unter anderen das „Bird 10,000 Genomes“ (*B10K*) - Projekt, welches sich zum Ziel gesetzt hat, die Genome aller 10.500 bekannten Vogelarten bis zum Jahre 2020 zu sequenzieren [263]. Die durch dieses Projekt generierten Daten bieten eine ideale Grundlage zur Überprüfung und Bestätigung der Ergebnisse dieser Arbeit bezüglich *stk33* innerhalb der Klasse *Aves*.

Ein derartiger Verlust von Genen stellt einen häufigen und wichtigen evolutionären Prozess dar und wurde sowohl in Bakterien [264] als auch in Eukaryoten [265] und Archaeen [266] nachgewiesen. In einigen Gruppen, wie z.B. Bakterien oder Tunikaten, welche im Verlauf der Evolution eine erheblich Reduzierung ihrer Genome durchgemacht haben [264; 238–241], ist solch ein Verlust von Genen besonders häufig zu beobachten. Auch in der Klasse *Aves* kann solch eine Reduzierung der Genomgröße und der damit verbundene Verlust von Genen im Vergleich zu anderen Tetrapoden-Klassen beobachtet werden, jedoch in einem deutlich geringerem Ausmaß als in Bakterien oder Tunikaten [267–269].

Interessanterweise konnte in allen Neukiefervögeln, neben der Deletion von Exon 3, keine andere Mutation nachgewiesen werden, die in allen untersuchten Vogelarten vorkommt (siehe Abbildung 3.11 & Abbildung A 9). Zwar wäre es auch denkbar, dass ein Indel oder eine Punktmutation, die ein vorzeitiges Stopkodon in den offenen Leserahmen einführen, in dem gemeinsamen Vorfahren aller Neukiefervögel ursächlich für den Verlust von *stk33* waren und im weiteren Verlauf der Evolution dann in verschiedenen Spezies erneut mutierten. Somit

wäre diese ursprüngliche Mutation nicht mehr direkt nachweisbar. Jedoch stellt die Theorie, dass die chromosomale Deletion von Exon 3 die initiale Mutation für die Pseudogenisierung von *stk33* war, hier die plausiblere Möglichkeit dar. Diese Mutation trat höchst wahrscheinlich in einer sehr frühen Phase der Diversifikation der Klasse *Aves* während der späten Kreidezeit vor ungefähr 100 Millionen Jahren auf, am Verzweigungspunkt von Neognathae und Palaeognathae (Abbildung A 8) [215–217]. Eine aktuelle Studie, die sich mit der vergleichenden Genomanalyse in Vögeln beschäftigt, konnte zeigen, dass die Entstehung der Neukiefervögeln mit einer erhöhten Rate an genomischen Rearrangements begleitet wurde [217]. Dies und die lange Zeitspanne seit dem Verlust von Exon 3 und der damit einhergehenden Pseudogenisierung von *stk33* stehen im Einklang mit den beobachteten drastischen Veränderungen innerhalb des *stk33*-Locus in Neukiefervögeln. Trotz alledem können solche Pseudogene durch die Expression von nicht-kodierenden RNAs (ncRNAs) neue Funktionen erlangen [270], jedoch konnten trotz intensiver Suchen keinerlei Transkripte des *stk33*-Pseudogens gefunden werden. Daher kann eine mögliche neu erlangte Funktion des *stk33* Pseudogens ausgeschlossen werden. Dies führt zu der Frage, welches Gen die Funktion von *stk33* übernommen hat, oder ob diese in Neukiefervögeln nicht mehr benötigt wird.

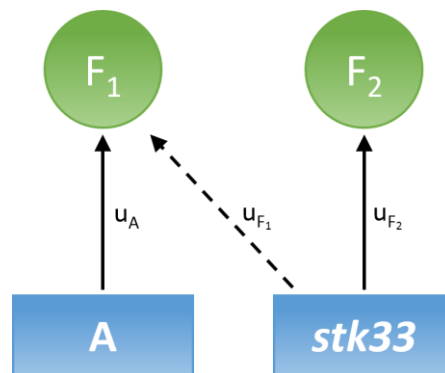
Im Folgenden sollen daher verschiedene Hypothesen zu dem Verlust von *stk33* und dessen Fixierung in Neognathae diskutiert werden. Die einfachste dieser Hypothesen geht von der Annahme aus, dass die Funktion von *stk33* in Neukiefervögeln überflüssig geworden ist und das Pseudogen so durch genetische Drift fixiert werden konnte [271], da der Verlust von *stk33* in diesem Fall keinen Nachteil für den Organismus darstellt und somit nicht mehr der negativen Selektion unterliegt. Dieser Mechanismus liegt mit hoher Wahrscheinlichkeit der voneinander unabhängigen Pseudogenisierung der L-Gulonolactonoxidase in verschiedenen Vertebraten wie den Knochenfischen (*Teleostei*), Primaten (*Primates*), Meerschweinchen (*Caviidae*) sowie in bestimmten Arten der Fledermäuse (*Microchiroptera*) und Sperlingsvögel (*Passeriformes*) zugrunde. Die L-Gulonolactonoxidase katalysiert durch die Oxidation von L-Gulonolacton in zahlreichen Vertebraten und auch Invertebraten den letzten Schritt der Ascorbinsäuresynthese (Vitamin C). Da diese Arten jedoch über ein Vitamin C reiches Nahrungsspektrum verfügen, sind diese Tiere nicht mehr auf die eigene Herstellung der Ascorbinsäure angewiesen, sondern können diese in ausreichendem Maße über die Nahrung aufnehmen. Dieses Angebot von Ascorbinsäure in der Nahrung hat dann zu einem Verlust der negativen Selektion für die beteiligten Gene und somit zur Pseudogenisierung dieser Gene geführt [272–277]. Für *stk33* ist dieses Szenario jedoch unwahrscheinlich, da *stk33* mit Funktionen von entscheidender Bedeutung, wie der Spermatogenese bzw. der Organogenese in Verbindung gebracht wird [20] und es keinerlei Hinweise für die Annahme gibt, dass

Neukiefervögel ohne die Funktion von *stk33* auskommen, während Urkiefervögel und alle anderen Vertebraten dies offensichtlich nicht können.

Eine zweite Möglichkeit ist durch die sogenannte „less-is-more“ Hypothese gegeben. Diese besagt, dass der Verlust eines Gens vorteilhaft sein kann und daher als Triebkraft für evolutionären Wandel dienen kann [278; 279]. Obwohl dies zunächst widersprüchlich erscheinen mag, da diese Hypothese evolutionäre Innovation durch Degeneration nahelegt, sind verschiedene Beispiele für solch einen vorteilhaften Verlust von Genen bekannt. So wird der Verlust der  $\alpha$ -1,3-Galactosyltransferase in Altweltaffen (*Catarrhini*) mit der Immunität gegen ein altertümliches Virus in Verbindung gebracht [280]. Weiterhin wird der Verlust der Gene „*CMP-N-acetylneuraminic acid hydroxylase*“ und „*myosin heavy chain*“ im Menschen mit der Rückbildung des Gebisses und der so ermöglichten Enzephalisation während der frühen Evolution der Gattung *Homo* assoziiert [281; 282]. Man könnte nun argumentieren, dass der Verlust von *stk33* vorteilhaft sei, da es in mehreren Studien als essentiell für das Überleben verschiedener Arten von Krebszellen identifiziert werden konnte [27; 28; 26; 283; 34–37], jedoch ist diese Rolle von *stk33* nach wie vor höchst umstritten [30–33] und es gibt außerdem bisher keine Hinweise für einen vorteilhaften Effekt durch den Verlust von *stk33*, so dass dies eine eher unwahrscheinliche Hypothese darstellt. Gleichwohl stellt die Fragestellung, ob Neukiefervögel irgendeine Form der Immunität gegenüber diesen Arten von Krebs aufweisen, einen interessanten Ansatz dar.

Eine weitere mögliche Erklärung für die Pseudogenisierung von *stk33* ist das Phänomen der genetischen Redundanz. Diese kann sowohl durch Genduplikationen, als auch durch konvergente evolutionäre Prozesse entstehen, die zu Proteinen führen, die die gleiche Funktion erfüllen aber mit Bezug auf Ihre Sequenz keinerlei Ähnlichkeiten aufweisen [284; 285]. Aufgrund des fehlenden Selektionsdrucks auf diese Gene wird die genetische Redundanz oft als evolutionär instabil angesehen, was zur Fixierung von Mutationen in einem der redundanten Gene führen würde. Jedoch findet sich solch eine genetische Redundanz recht häufig in den Genomen höherer Organismen, was impliziert, dass redundante Gene unter bestimmten Bedingungen evolutionär stabil sein können [285; 286]. So konnte für die Protein Kinasen C delta und epsilon eine redundante Rolle während der Embryonalentwicklung der Maus nachgewiesen werden [287], die Callose-Synthasen GSL1 und GSL5 haben eine redundante Funktion in der Pollenentwicklung und Fertilität [288] und die Gene *vav1* und *vav3* spielen eine redundante Rolle bei der Aktivierung der Thrombozyten [289]. Verschiedene Modelle für solch eine stabile genetische Redundanz sind publiziert worden. Eines dieser Modelle berücksichtigt das Phänomen, dass ein Gen mehrere unterschiedliche Funktionen haben kann (Pleiotropie). Wenn das Gen A die Funktion  $F_1$  und Gen B sowohl die Funktion  $F_1$

als auch eine zweite Funktion  $F_2$  ausführt, Funktion  $F_1$  jedoch nur mit einer leicht reduzierten Effizienz, so wird diese genetische Redundanz als stabil angesehen. Dies gilt jedoch nur, sofern die Mutationsrate  $u_{F_1}$ , welche die redundante Funktion von *stk33* eliminiert, kleiner als die Mutationsrate  $u_A$ , welche die Funktion A inaktiviert, ist (siehe Abbildung 4.2) [290]. Obwohl die genaue Funktion von STK33 nach wie vor unklar ist, so deutet das Vorhandensein mehrerer alternativ gespleißter Transkriptvarianten in Mensch und Maus [9] sowie im Vogelstrauß (siehe 3.3.1) darauf hin, dass auch STK33 mehrere Funktionen übernimmt und das Vimentin wahrscheinlich nicht das einzige Substrat darstellt. Dies wird durch die Beobachtung bestätigt, dass STK33 auch in Zellen vorkommt, in denen kein Vimentin nachweisbar ist [24]. Setzt man für dieses Modell *stk33* mit Gen B gleich und nimmt gleichzeitig an, dass die Funktion  $F_2$  für Neukiefervögel überflüssig geworden ist, so hätte dies aufgrund des wegfallenden Selektionsdrucks zwangsläufig zu dem beobachteten Verlust von *stk33* geführt.



**Abbildung 4.2 | Modell einer stabilen genetischen Redundanz.** Wenn Gen A und *stk33* die unterschiedlichen Funktionen  $F_1$  bzw.  $F_2$  ausführen, *stk33* aber zusätzlich noch die Funktion  $F_1$  mit leicht reduzierter Effizienz ausführt, dann ist diese redundante Organisation stabil. Dies gilt jedoch nur, sofern die Mutationsrate  $u_{F_1}$ , welche die redundante Funktion für  $F_1$  von *stk33* eliminiert, kleiner als die Mutationsrate  $u_A$  ist. (verändert nach: Nowak, MA; Boerlijst, MC; Cooke, J; Smith, JM: *Evolution of genetic redundancy*. *Nature* 1997, 6638:167–171)

Eine Variante letzterer Hypothese ist die Möglichkeit, dass STK33 durch eine nicht-orthologe, aber funktional analoge Kinase ersetzt worden ist („*non-orthologous gene displacement*“, NOD) und somit überflüssig wurde. Auch in diesem Fall hätte solch ein Ereignis den Verlust des Selektionsdrucks und somit auch den Verlust von *stk33* zur Folge. Verschiedene Beispiele für solch ein NOD wurden bereits beschrieben. So kommt die Fructose-1,6-bisphosphate Aldolase, die eine essentielle Rolle sowohl in der Glykolyse als auch im Calvin-Zyklus spielt, in zwei verschiedenen Formen in vielzelligen Eukaryoten einerseits und in Bakterien und Hefen andererseits vor [291–293]. Diese beiden Formen der Fructose-1,6-bisphosphate Aldolase zeigen keinerlei Ähnlichkeit in Bezug auf ihre Sequenz und Struktur, katalysieren jedoch die gleiche Reaktion [292; 294]. Die Phosphomevalonat-Kinase, Bestandteil der Isoprenoid-Biosynthese, wird durch die zwei nicht-orthologen Gene *erg8* und *pmk* kodiert. Während *erg8*

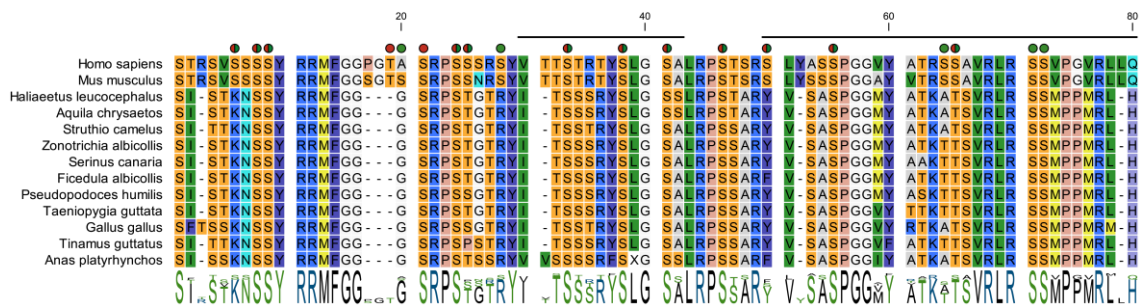
in Bakterien, Pilzen und Pflanzen vorkommt, kann *pmk* ausschließlich in Tieren gefunden werden, was auf ein NOD während der frühen Evolution der Tiere (*Animalia*) hindeutet [295]. Unabhängig davon, ob *stk33* aufgrund von genetischer Redundanz oder eines NOD in Neukiefervögeln verloren gegangen ist, muss eine weitere Kinase, die in der Lage ist STK33 zu phosphorylieren, beteiligt gewesen sein, wenn man davon ausgeht, dass die Funktion von STK33 nicht einfach überflüssig geworden ist. Die Funktion dieser Kinase ist entweder redundant zu der von STK33 mit Bezug auf die Phosphorylierung von Vimentin oder sie hat STK33 funktional ersetzt. Hierfür kommen nur Kinasen als Kandidaten in Betracht, die in der Lage sind Vimentin in der gleichen Region wie STK33 zu phosphorylieren. Zwar sind die genauen Phosphorylierungsstellen von STK33 bisher noch nicht bekannt, konnten aber mit Hilfe von Kinase-Assays mit verkürzten Vimentin Mutanten auf die amino-terminale nicht- $\alpha$ -helikale Kopfdomäne, die sich von Aminosäure 1 bis 80 erstreckt, eingegrenzt werden. Die Ergebnisse dieser Versuche sind indikativ für mindestens eine Phosphorylierungsstelle zwischen den Aminosäuren 30 und 42, sowie mindestens eine weitere zwischen den Aminosäuren 50 und 80 [24]. Insgesamt wurden bisher zehn Kinasen beschrieben, die in der Lage sind Vimentin zu phosphorylieren (siehe Tabelle 4.1). Alle diese Kinasen sind in der Klasse *Aves* vorhanden [267]. Aufgrund der Möglichkeit, dass Vimentin in Neukiefervögeln oder in Vögeln generell auf andere Weise reguliert werden könnte, wurde ein Alignment mit der amino-terminalen Kopfdomäne aller vollständig verfügbaren Vimentin Proteine aus der Klasse *Aves* zusammen mit den Vimentin Proteinsequenzen von Mensch und Maus erstellt. Hierbei zeigt sich, dass die meisten Phosphorylierungsstellen zwischen Mensch, Maus und den verschiedenen Vogelarten, insbesondere in den Bereichen, in denen STK33 phosphoryliert, konserviert ist (Abbildung 4.3). Dies deutet darauf hin, dass Vimentin in Vögeln auf ähnlichem Wege wie in Mensch und Maus reguliert wird.

Die Phosphorylierungsstellen zweier Kinasen, Plk1 und p37, befinden sich außerhalb der amino-terminalen Kopfdomäne [296; 297] und können daher als Kandidaten für die Substitution von STK33 ausgeschlossen werden. Drei weitere Kinasen (Pak1, CDK1, Aurora Kinase B) stellen ebenfalls keine geeigneten Kandidaten dar, da sie Vimentin an nur einer einzelnen Stelle phosphorylieren [298–303]. Obwohl MAPKAP Kinase 2 und CAMKII mindestens eine Phosphorylierungsstelle innerhalb derselben Region wie STK33 haben, phosphorylieren sie auch Ser-82 [304–307] und sind somit eher unwahrscheinliche Substituenten. Die Phosphorylierungsstellen der drei verbleibenden Kinasen (Rho-Kinase, Protein Kinase C und Protein Kinase A) liegen ausnahmslos innerhalb derselben Region, die auch von STK33 phosphoryliert wird (Tabelle 4.1), jedoch zeigt das Vimentin-Alignment, dass das durch die Protein Kinase C phosphorylierte Ser-50 [308; 309] in der Klasse *Aves* nicht

konserviert ist. Hier wurde das Serin durch ein Tyrosin, welches durch Serin/Threonin-Kinasen nicht phosphoryliert werden kann, substituiert (vgl. Abbildung 4.3). Daraus lässt sich die Schlussfolgerung ziehen, dass Protein Kinase A und Rho-Kinase die vielversprechendsten Kandidaten für eine funktionelle Substitution von STK33 darstellen, da ihre Phosphorylierungsstellen innerhalb derselben Region wie die von STK33 liegen und in der Klasse *Aves* hoch konserviert sind. Um hierüber eine genauere Aussage treffen zu können sind jedoch weitere Untersuchungen bezüglich der genauen Funktion von STK33 sowie dessen exakten Phosphorylierungsstellen notwendig.

**Tabelle 4.1 | Bekannte Vimentin-phosphorylierende Kinasen und deren Phosphorylierungsstellen.**

Kinase	Phosphorylierungsstelle(n)	Literatur
Rho-Kinase	Ser-38; Ser-71	[310]
Protein Kinase C	Ser-33; Ser-50	[308; 309]
CaMKII	Ser-38; Ser-82	[304; 305]
Protein Kinase A	Ser-38; Ser-72	[311]
MAPKAP Kinase-2	Ser-38; Ser-50; Ser-55; Ser-82	[306; 307]
Aurora Kinase B	Ser-72	[298]
CDK1	Ser-55	[299–301]
Pak1	Ser-55	[302; 303]
Plk1	Ser-82	[296]
p37	Ser-457; Ser-458	[297]



**Abbildung 4.3 | Alignment der amino-terminalen Kopfdomäne von Vimentin aus verschiedenen Vogelspezies mit Mensch und Maus.** Die Phosphorylierungsstellen des Menschen [312; 313] sind durch einen roten Punkt, die Phosphorylierungsstellen der Maus [314] sind durch einen grünen Punkt markiert. Phosphorylierungsstellen, welche sowohl im Menschen als auch in der Maus vorkommen sind durch je einen roten und grünen Halbkreis gekennzeichnet. Die Bereiche, in denen die Phosphorylierungsstellen von STK33 liegen, sind durch horizontale, schwarze Balken markiert. Dargestellt sind die amino-terminalen Kopfdomänen der Vimentin Proteine von Mensch (*Hsa*), Maus (*Mmu*), Weißkopfschwarzadler (*Hle*), Steinadler (*Ach*), Vogelstrauß (*Sca*), Weißkehlammer (*Zal*), Kanariengirlitz (*Sca*), Halsbandschnäpper (*Fal*), Tibetmeise (*Phu*), Zebrafink (*Tae*), Haushuhn (*Gga*), Weißkehlitnamu (*Tgu*) und Stockente (*Apl*). Es zeigt sich deutlich, dass der Großteil der Phosphorylierungsstellen aus Mensch und Maus auch in der Klasse *Aves* konserviert sind.



# 5 | Zusammenfassung

Die Serin/Threonin-Kinase 33 (*stk33*) konnte im Zuge einer vergleichenden Genomanalyse des humanen Chromosomenabschnittes 11p15.3 sowie dessen orthologer Region in der Maus [9; 14; 15] identifiziert werden. Weiterführende Analysen konnten zeigen, dass *stk33* vornehmlich in Testes, bestimmten Hirnregionen und embryonalen Organen exprimiert wird, was zu der Hypothese geführt hat, dass *stk33* in Spermatogenese und Organogenese involviert ist [20]. Das STK33-Protein konnte als Mitglied der Calcium/Calmodulin abhängigen Kinasen klassifiziert werden [9] und es konnte gezeigt werden, dass STK33 mit Vimentin ko-lokalisiert [23] und dieses phosphoryliert [24]. Weiterhin wurde STK33 mit verschiedenen Arten von Krebs in Verbindung gebracht [26–28; 283; 34–37], jedoch ist diese Rolle nicht unumstritten [30–33]. Bisher konnte *stk33* in allen Vertebraten, mit Ausnahme der Vögel, nachgewiesen werden.

In der vorliegenden Arbeit sollten zwei verschiedene Fragestellungen bearbeitet werden. Zum einen sollte aufgrund der bisher nach wie vor noch nicht vollständig bekannten physiologischen Funktion von STK33 ein Zebrafisch „*Knock-out*“ Stamm für *stk33* zunächst mit Hilfe der TALEN-Technologie und später mit Hilfe des CRISPR/Cas-Systems etabliert werden. Zum anderen sollte analysiert werden, ob *stk33* in der Klasse *Aves* tatsächlich nicht vorkommt und eventuell auch außerhalb des Unterstammes der Wirbeltiere (*Vertebrata*) vorkommt.

Zu Beginn dieser Arbeit wurde zunächst die gerade neu verfügbar gewordene TALEN-Technologie [194] für die Herstellung eines *stk33* „*Knock-out*“ Stammes verwendet. Hierbei kam es jedoch zu verschiedenen Problemen bei der Herstellung der TALEN-Vektoren und es konnten nach erfolgter Mikroinjektion keine Mutationen detektiert werden. Aufgrund dessen wurde anstatt dessen im weiteren Verlauf das ebenfalls neu verfügbar gewordene CRISPR/Cas-System [53] verwendet. Aufgrund der vorangegangenen Probleme mit den TALENs sollte zunächst das CRISPR/Cas-System sowie der gesamte Arbeitsablauf von der Herstellung und Mikroinjektion der unterschiedlichen RNAs bis hin zu der Genotypisierung der injizierten Embryonen anhand einer Positivkontrolle etabliert werden. Hierbei konnte das Gen *no tail (ntl)* erfolgreich ausgeschaltet und der entsprechende Phänotyp beobachtet werden, der sich durch das vollständige Fehlen der kaudalen Strukturen auszeichnet. Mit Hilfe des zuvor etablierten CRISPR/Cas-Systems konnte auch *stk33* erfolgreich in injizierten Embryonen ausgeschaltet werden. Bei der Aufzucht dieser Embryonen ist jedoch ein Großteil noch während der späteren

Larvalstadien verstorben, so dass nur ein kleiner Teil der injizierten Embryonen die Pubeszenz erreichte. Um die überlebenden Fische zu genotypisieren wurden diese ausgekreuzt und die daraus resultierende F<sub>1</sub>-Generation genotypisiert. Hierbei zeigten sich jedoch keinerlei Hinweise auf Mutationen innerhalb der *stk33* Zielregion. Auch bei der vollständigen Genotypisierung der adulten F<sub>0</sub>-Tiere konnten, trotz der positiven Genotypisierungen nach erfolgter Injektion, keine Mutationen nachgewiesen werden. Diese Ergebnisse führten zu der Hypothese, dass alle mutanten Embryonen noch während der späteren Larvalstadien aufgrund der potenziellen Rolle von *stk33* bei der Organogenese [20] verstorben sind und nur nicht mutante Embryonen bis zur Pubeszenz heranwachsen konnten. Das gewebspezifische Ausschalten von *stk33* in den Gonaden [235] könnte dieses Problem umgehen.

In dem zweiten Teil der vorliegenden Arbeit sollte ein eventuelles Vorhandensein von *stk33* außerhalb der Wirbeltiere (*Vertebrata*) sowie innerhalb der Vögel (*Aves*) analysiert werden. Hierzu wurden verschiedene Analysepipelines für die Analyse genomischer und transkriptomischer Datensätze etabliert. Auf diese Weise konnten insgesamt 726 genomische und transkriptomische Datensätze von 488 unterschiedlichen Tierarten auf ein Vorhandensein von *stk33* hin analysiert werden. Hierbei zeigte sich, dass *stk33* in den meisten Vertretern der vielzelligen Tiere (*Metazoa*) vorkommt und somit in dem gemeinsamen Vorfahren aller vielzelligen Tiere vor ca. 1,5 Milliarden Jahren entstanden sein muss. Es zeigte sich jedoch auch, dass *stk33* in mehreren Taxa nicht nachweisbar war, woraus sich schließen lässt, dass *stk33* mehrmals im Laufe der Evolution der *Metazoa* verloren gegangen sein muss. Innerhalb der Klasse *Aves* konnte *stk33* erfolgreich innerhalb der basalen Urkiefervögel (*Palaeognathae*) nachgewiesen werden. Die Analysen der Neukiefervögel konnten jedoch lediglich nicht funktionale Überreste von *stk33* nachweisen. Der genomische *stk33* Locus dieser Vögel ist von großen genomischen Deletionen sowie zahlreichen Indels und Stopkodons innerhalb der verbliebenen Exons geprägt. Daraus lässt sich schließen, dass *stk33* hier nur noch als unitäres Pseudogen vorliegt und bei der Abspaltung der Neukiefervögel während der späten Kreidezeit vor ca. 100 Millionen Jahren verloren gegangen sein muss. Da die Neukiefervögel sowie einige weitere Taxa trotz der möglichen Beteiligungen von *stk33* an solch wichtigen Funktionen wie der Spermatogenese oder der Organogenese offensichtlich ohne *stk33* auskommen, während die Urkiefervögel sowie alle anderen Tierarten dies offensichtlich nicht können, wird vermutet, dass die Pseudogenisierung von *stk33* entweder aufgrund einer genetischen Redundanz oder eines „*non-orthologous gene displacement*“ (NOD) fixiert werden konnte. Anhand der Analyse der Phosphorylierungsstellen für Vimentin konnten die Rho-Kinase und die Protein Kinase A als mögliche Kandidaten hierfür identifiziert werden.

# Abkürzungsverzeichnis

°C	Grad Celsius
µg	Mikrogramm
µl	Mikroliter
A.bidest.	Aqua bidestillata
abs.	absolut
AIDS	acquired immunodeficiency syndrome
AS	Aminosäure
Asx	Asparagin und Asparaginsäure
ATP	Adenosintriphosphat
BAC	bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
BLASTn	Basic Local Alignment Search Tool – Nukleotidsequenz gg. Nukleotiddatenbank
BLASTx	Basic Local Alignment Search Tool – translatierte Nukleotidsequenz gg. Proteindatenbank
bp	Basenpaare
Ca <sup>2+</sup>	Calcium-Ion
CAMK	Calcium/Calmodulin-abhängige Proteinkinase
CAMKII	Calcium/Calmodulin-abhängige Proteinkinase II
Cas	CRISPR-assoziierte Gene
cDNA	komplementäre DNA
CMV	Cytomegalievirus
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	crisprRNA
DNA	Desoxyribonukleinsäure
dNTP	2'-Desoxynukleosidtriphosphat
DSB	Doppelstrangbruch
DTT	Dithiothreitol
EDTA	Ethylendinitrilotetraessigsäure
E. coli	<i>Escherichia coli</i>
ESTs	expressed sequence tags
EZRC	European Zebrafish Ressource Center
FLASH	Fast Ligation-based Automatabale Solid-phase Hight-throughput
FU	fluorescence units
EtOH	Ethanol
g	Gramm

---

ges.	gesättigt
h	Stunde
HDR	Homology directed repair
hpf	hours post fertilisation
ICA	Iterative Capped Assembly
Indel	Insertion/Deletion
IPTG	Isopropyl- $\beta$ -D-Thiogalactosid
kb	Kilobasenpaare
kDa	Kilodalton
KDD	Knowledge discovery in databases
KIT	Karlsruhe Institute of Technology
konz.	konzentriert
lr-PCR	long range PCR
m	Milli
M	Molar
mA	Milli-Ampere
mRNA	messenger-RNA
miRNA	micro-RNA
min.	Minuten
MOPS	3(N-morpholin)PropanSulfonsäure
n	Nano
NCBI	National Center for Biotechnology Information
NGS	Next-Generation Sequencing
NHEJ	Non-homologous end joining
NOD	non-orthologous gene displacement
ntl	Gen no-tail
ORF	open reading frame
PAM	protospacer adjacent motif
PCR	Polymerase chain reaction“
psi	Pound-force per square inch
qRT-PCR	quantitative real-Time Polymerase chain reaction
RACE-PCR	rapid amplification of cDNA-ends polymerase chain reaction
REAL	Restriction Enzyme and Ligation
RNA	Ribonukleinsäure
RNAi	RNA interference
RT	Reverse Transkriptase
RT-PCR	reverse transcriptase polymerase chain reaction
RVD	repeat variable diresidue
s	Sekunde
SCID	severe combined immunodeficiency syndrome

---

SDS	Natriumdodecylsulfat
Ser-	Serin
sgRNA	single guide RNA
siRNA	small interfering RNA
SMRT	single molecule real time
SNP	Single Nucleotide Polymorphism
SRA	Sequence Read Archive
ssDNA	einzelsträngige DNA
<i>stk33</i>	Gen für die Serin/Threonin-Kinase 33
STK33	Proteinprotein der Serin/Threonin-Kinase 33
T7E1	T7 Endonuklease I
TALEs	Transcription activator-like effectors
TALENs	Transcription activator-like effector nucleases
Taq	<i>Thermus aquaticus</i>
TBE	Tris-Borat-EDTA-Puffer
TFO	triplex-forming oligonucleotides
tracrRNA	trans-activating crisperRNA
Tris	Tris(hydroxymethyl)aminomethan
tru-sgRNA	truncated single guide RNA
U	unit
U/min	Umdrehungen pro Minute
UTR	untranslatierte Region
UV	Ultraviolettes Licht
VE-Wasser	Voll entsalztes Wasser
Vol.	Volumen
w/v	weight per volume
ZFN	Zinkfingernukleasen
ZFP	Zinkfingerprotein

# Literaturverzeichnis

- [1] Adams, J. A. (2001). Kinetic and catalytic mechanisms of protein kinases. *Chem. Rev.* 101 (8), 2271–2290.
- [2] Berg, Jeremy M.; Tymoczko, John L.; Stryer, Lubert (2003). *Biochemie*. 5. Aufl. Heidelberg [u.a.]: Spektrum Akad. Verl.
- [3] Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science* 298 (5600), 1912–1934.
- [4] Manning, Gerard (2005). Genomic overview of protein kinases. *WormBook*, 1–19.
- [5] Krauss, Gerhard (2003). *Biochemistry of signal transduction and regulation*. 3. Aufl. Weinheim: Wiley-VCH.
- [6] Hanks, Steven K. (2003). Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biol.* 4 (5), 111.
- [7] Hanks, S. K.; Hunter, T. (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* 9 (8), 576–596.
- [8] Capra, M. (2006). Frequent Alterations in the Expression of Serine/Threonine Kinases in Human Cancers. *Cancer Research* 66 (16), 8147–8154.
- [9] Mujica, A. O.; Hankeln, T.; Schmidt, E. R. (2001). A novel serine/threonine kinase gene, STK33, on human chromosome 11p15.3. *Gene* 280 (1-2), 175–181.
- [10] Swulius, M. T.; Waxham, M. N. (2008). Ca(2+)/calmodulin-dependent protein kinases. *Cell. Mol. Life Sci.* 65 (17), 2637–2657.
- [11] Wilmann, M.; Gautel, M.; Mayans, O. (2000). Activation of calcium/calmodulin regulated kinases. *Cellular and molecular biology (Noisy-le-Grand, France)* 46 (5), 883–894.
- [12] Griffith, Leslie C. (2004). Regulation of calcium/calmodulin-dependent protein kinase II activation by intramolecular and intermolecular interactions. *J. Neurosci.* 24 (39), 8394–8398.
- [13] Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J. et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409 (6822), 860–921.
- [14] Amid, C.; Bahr, A.; Mujica, A.; Sampson, N.; Bikar, S. E.; Winterpacht, A. et al. (2001). Comparative genomic sequencing reveals a strikingly similar architecture of a conserved syntenic region on human chromosome 11p15.3 (including gene ST5) and mouse chromosome 7. *Cytogenet. Cell Genet.* 93 (3-4), 284–290.
- [15] Cichutek, A.; Brueckmann, T.; Seipel, B.; Hauser, H.; Schlaubitz, S.; Prawitt, D. et al. (2001). Comparative architectural aspects of regions of conserved synteny on human chromosome 11p15.3 and mouse chromosome 7 (including genes WEE1 and LMO1). *Cytogenet Cell Genet* 93 (3-4), 277–283.
- [16] Nowak, Norma J.; Shows, Thomas B. (1995). Genetics of Chromosome 11: Loci for Pediatric and Adult Malignancies, Developmental Disorders, and Other Diseases. *Cancer Invest* 13 (6), 646–659.
- [17] Karnik, P.; Paris, M.; Williams, B. R.; Casey, G.; Crowe, J.; Chen, P. (1998). Two distinct tumor suppressor loci within chromosome 11p15 implicated in breast cancer progression and metastasis. *Human molecular genetics* 7 (5), 895–903.
- [18] Schwartz, Benedikt (2006). Charakterisierung und Funktionsanalyse des Promotors des STK/Stk 33-Gens bei Mensch und Maus. Diplomarbeit. Johannes Gutenberg Universität Mainz, Mainz. Institut für Molekulargenetik.
- [19] Mujica, Alejandro O. (2004). Comparative genomic sequencing analysis of a region in human chromosome 11p15.3/mouse chromosome 7 and analysis of the novel STK33/Stk33 gene. Dissertation. Johannes Gutenberg Universität Mainz, Mainz. Institut für Molekulargenetik.
- [20] Mujica, Alejandro O.; Brauksiepe, Bastienne; Saaler-Reinhardt, Sigrid; Reuss, Stefan; Schmidt, Erwin R. (2005). Differential expression pattern of the novel serine/threonine kinase, STK33, in mice and men. *FEBS J.* 272 (19), 4884–4898.
- [21] Janca, F. C.; Jost, L. K.; Evenson, D. P. (1986). Mouse testicular and sperm cell development characterized from birth to adulthood by dual parameter flow cytometry. *Biology of reproduction* 34 (4), 613–623.
- [22] Kaynak, Bogac; Heydebreck, Anja von; Mebus, Siegrun; Seelow, Dominik; Hennig, Steffen; Vogel, Jan et al. (2003). Genome-wide array analysis of normal and malformed human hearts. *Circulation* 107 (19), 2467–2474.
- [23] Brauksiepe, Bastienne; Baumgarten, Lisa; Reuss, Stefan; Schmidt, Erwin R. (2013). Co-localization of serine/threonine kinase 33 (Stk33) and vimentin in the hypothalamus. *Cell Tissue Res* 355 (1), 189–199.
- [24] Brauksiepe, Bastienne; Mujica, Alejandro O.; Herrmann, Harald; Schmidt, Erwin R. (2008). The Serine/threonine kinase Stk33 exhibits autophosphorylation and phosphorylates the intermediate filament protein Vimentin. *BMC Biochem.* 9, 25.
- [25] Reuss, Stefan; Brauksiepe, Bastienne; Disque-Kaiser, Ursula; Olivier, Tim (2017). Serine/threonine-

- kinase 33 (Stk33) – Component of the neuroendocrine network? *Brain Res.* 1655, 152–160.
- [26] Frohling, S.; Scholl, C. (2011). STK33 Kinase Is Not Essential in KRAS-Dependent Cells-Letter. *Cancer Research* 71 (24), 7716.
- [27] Scholl, Claudia; Fröhling, Stefan; Dunn, Ian F.; Schinzel, Anna C.; Barbie, David A.; Kim, So Young et al. (2009). Synthetic Lethal Interaction between Oncogenic KRAS Dependency and STK33 Suppression in Human Cancer Cells. *Cell* 137 (5), 821–834.
- [28] Azoitei, Ninel; Hoffmann, Christopher M.; Ellegast, Jana M.; Ball, Claudia R.; Obermayer, Kerstin; Gößele, Ulrike et al. (2012). Targeting of KRAS mutant tumors by HSP90 inhibitors involves degradation of STK33. *J. Exp. Med.* 209 (4), 697–711.
- [29] Downward, Julian (2009). Finding the weakness in cancer. *N. Engl. J. Med.* 361 (9), 922–924.
- [30] Babij, C.; Zhang, Y.; Kurzeja, R. J.; Munzli, A.; Shehabeldin, A.; Fernando, M. et al. (2011). STK33 Kinase Activity Is Nonessential in KRAS-Dependent Cancer Cells. *Cancer Research* 71 (17), 5818–5826.
- [31] Luo, T.; Masson, K.; Jaffe, J. D.; Silkworth, W.; Ross, N. T.; Scherer, C. A. et al. (2012). STK33 kinase inhibitor BRD-8899 has no effect on KRAS-dependent cancer cell viability. *Proceedings of the National Academy of Sciences* 109 (8), 2860–2865.
- [32] Weiwer, Michel; Spoonamore, James; Wei, Jingqiang; Guichard, Boris; Ross, Nathan T.; Masson, Kristina et al. (2012). A Potent and Selective Quinoxalinone-Based STK33 Inhibitor Does Not Show Synthetic Lethality in KRAS-Dependent Cells. *ACS Med Chem Lett* 3 (12), 1034–1038.
- [33] Spoonamore J, Weiwer M, Wei J, et al. (2011 Dec 16 [Updated 2014 May 13]). Screen for Inhibitors of STK33 Kinase Activity. Hg. v. Probe Reports from the NIH Molecular Libraries Program [Internet]. Online verfügbar unter <http://www.ncbi.nlm.nih.gov/books/NBK133418/>.
- [34] Moon, Ji Wook; Lee, Soo Kyung; Lee, Jung Ok; Kim, Nami; Lee, Yong Woo; Kim, Su Jin et al. (2014). Identification of novel hypermethylated genes and demethylating effect of vincristine in colorectal cancer. *J. Exp. Clin. Cancer Res.* 33, 4.
- [35] Yang, Tian; Song, Bin; Zhang, Jin; Yang, Guang-Shun; Zhang, Han; Yu, Wei-Feng et al. (2014). STK33 promotes hepatocellular carcinoma through binding to c-Myc. *Gut.*
- [36] Huang, Lingyan; Chen, Chen; Zhang, Guodong; Ju, Yuanrong; Zhang, Jianzhong; Wang, Haibo; Li, Jianfeng (2015). STK33 overexpression in hypopharyngeal squamous cell carcinoma: possible role in tumorigenesis. *BMC cancer* 15 (1), 13.
- [37] Wang, Ping; Cheng, Hongzhong; Wu, Jianqiang; Yan, Anrun; Zhang, Libin (2015). STK33 plays an important positive role in the development of human large cell lung cancers with variable metastatic potential. *Acta Biochim. Biophys. Sin. (Shanghai)* 47 (3), 214–223.
- [38] Piccaluga, Pier Paolo; Navari, Mohsen; Falco, Giulia De; Ambrosio, Maria Raffaella; Lazzi, Stefano; Fuligni, Fabio et al. (2015). Virus-encoded microRNA contributes to the molecular profile of EBV-positive burkitt lymphomas. *Oncotarget.*
- [39] Nobel Media AB 2014. The Nobel Prize in Physiology or Medicine 2007. Web. 29 Jun 2015. Nobelprize.org. Online verfügbar unter [www.nobelprize.org/nobel\\_prizes/medicine/laureates/2007/](http://www.nobelprize.org/nobel_prizes/medicine/laureates/2007/), zuletzt geprüft am 29.06.2015.
- [40] Kim, Hyongbum; Kim, Jin-Soo (2014). A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* 15 (5), 321–334.
- [41] Kolb, Andreas F. (2002). Genome engineering using site-specific recombinases. *Cloning and stem cells* 4 (1), 65–80.
- [42] Ivics, Zoltán; Li, Meng Amy; Mátés, Lajos; Boeke, Jef D.; Nagy, Andras; Bradley, Allan; Izsvák, Zsuzsanna (2009). Transposon-mediated genome manipulation in vertebrates. *Nat Meth* 6 (6), 415–422.
- [43] Lieber, Michael R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 79, 181–211.
- [44] Urnov, Fyodor D.; Miller, Jeffrey C.; Lee, Ya-Li; Beausejour, Christian M.; Rock, Jeremy M.; Augustus, Sheldon et al. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435 (7042), 646–651.
- [45] Lee, Hyung Joo; Kim, Eunji; Kim, Jin-Soo (2010). Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res.* 20 (1), 81–89.
- [46] Lee, Hyung Joo; Kweon, Jiyeon; Kim, Eunji; Kim, Seokjoong; Kim, Jin-Soo (2012). Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res.* 22 (3), 539–548.
- [47] Brunet, Erika; Simsek, Deniz; Tomishima, Mark; DeKolver, Russell; Choi, Vivian M.; Gregory, Philip et al. (2009). Chromosomal translocations induced at specified loci in human stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 106 (26), 10620–10625.
- [48] Kim, Young-Kook; Wee, Gabbine; Park, Joha; Kim, Jongkyu; Baek, Daehyun; Kim, Jin-Soo; Kim, V. Narry (2013). TALEN-based knockout library for human microRNAs. *Nature structural & molecular biology* 20 (12), 1458–1464.
- [49] Hu, Ruozhen; Wallace, Jared; Dahlem, Timothy J.; Grunwald, David Jonah; O'Connell, Ryan M. (2013). Targeting human microRNA genes using engineered Tal-effector nucleases (TALENs). *PLoS ONE* 8 (5), e63074.
- [50] Gutschner, Tony; Baas, Marion; Diederichs, Sven (2011). Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. *Genome Res.* 21 (11), 1944–1954.

- [51] Zou, Jizhong; Mali, Prashant; Huang, Xiaosong; Dowey, Sarah N.; Cheng, Linzhao (2011). Site-specific gene correction of a point mutation in human iPSCs derived from an adult patient with sickle cell disease. *Blood* 118 (17), 4599–4608.
- [52] Chen, Fuqiang; Pruett-Miller, Shondra M.; Huang, Yuping; Gjoka, Monika; Duda, Katarzyna; Taunton, Jack et al. (2011). High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Meth* 8 (9), 753–755.
- [53] Hwang, Woong Y.; Fu, Yanfang; Reyon, Deepak; Maeder, Morgan L.; Tsai, Shengdar Q.; Sander, Jeffrey D. et al. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 31 (3), 227–229.
- [54] Wefers, Benedikt; Meyer, Melanie; Ortiz, Oskar; Hrabé de Angelis, Martin; Hansen, Jens; Wurst, Wolfgang; Kühn, Ralf (2013). Direct production of mouse disease models by embryo microinjection of TALENs and oligodeoxynucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 110 (10), 3782–3787.
- [55] Cui, Xiaoxia; Ji, Diana; Fisher, Daniel A.; Wu, Yumei; Briner, David M.; Weinstein, Edward J. (2011). Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nat Biotechnol* 29 (1), 64–67.
- [56] Sung, Young Hoon; Baek, In-Jeoung; Kim, Duk Hyoung; Jeon, Jisun; Lee, Jaehoon; Lee, Kyunghee et al. (2013). Knockout mice created by TALEN-mediated gene targeting. *Nat Biotechnol* 31 (1), 23–24.
- [57] Li, Dali; Qiu, Zhongwei; Shao, Yanjiao; Chen, Yuting; Guan, Yuting; Liu, Meizhen et al. (2013). Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat Biotechnol* 31 (8), 681–683.
- [58] Zou, Jizhong; Maeder, Morgan L.; Mali, Prashant; Pruett-Miller, Shondra M.; Thibodeau-Beganny, Stacey; Chou, Bin-Kuan et al. (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell stem cell* 5 (1), 97–110.
- [59] Shukla, Vipula K.; Doyon, Yannick; Miller, Jeffrey C.; DeKolver, Russell C.; Moehle, Erica A.; Worden, Sarah E. et al. (2009). Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459 (7245), 437–441.
- [60] Townsend, Jeffrey A.; Wright, David A.; Winfrey, Ronnie J.; Fu, Fengli; Maeder, Morgan L.; Joung, J. Keith; Voytas, Daniel F. (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* 459 (7245), 442–445.
- [61] Li, Ting; Liu, Bo; Spalding, Martin H.; Weeks, Donald P.; Yang, Bing (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol* 30 (5), 390–392.
- [62] Hauschild, Janet; Petersen, Bjoern; Santiago, Yolanda; Queisser, Anna-Lisa; Carnwath, Joseph W.; Lucas-Hahn, Andrea et al. (2011). Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* 108 (29), 12013–12017.
- [63] Carlson, Daniel F.; Tan, Wenfang; Lillico, Simon G.; Stverakova, Dana; Proudfoot, Chris; Christian, Michelle et al. (2012). Efficient TALEN-mediated gene knockout in livestock. *Proc. Natl. Acad. Sci. U.S.A.* 109 (43), 17382–17387.
- [64] Yu, Shengli; Luo, Junjie; Song, Zhiyuan; Ding, Fangrong; Dai, Yunping; Li, Ning (2011). Highly efficient modification of beta-lactoglobulin (BLG) gene via zinc-finger nucleases in cattle. *Cell research* 21 (11), 1638–1640.
- [65] Sanjukta Misra (2013). Human Gene Therapy : A Brief Overview of the Genetic Revolution. *J Assoc Physicians India* (61), 127–133.
- [66] Cavazzana-Calvo, M. (2000). Gene Therapy of Human Severe Combined Immunodeficiency (SCID)-X1 Disease. *Science* 288 (5466), 669–672.
- [67] Check, Erika (2002). Gene therapy: A tragic setback. *Nature* 420 (6912), 116–118.
- [68] Parekh-Olmedo, H.; Ferrara, L.; Brachman, E.; Kmiec, E. B. (2005). Gene therapy progress and prospects: targeted gene repair. *Gene Ther* 12 (8), 639–646.
- [69] Niu, Yuyu; Shen, Bin; Cui, Yiqiang; Chen, Yongchang; Wang, Jianying; Wang, Lei et al. (2014). Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* 156 (4), 836–843.
- [70] Liang, Puping; Xu, Yanwen; Zhang, Xiya; Ding, Chenhui; Huang, Rui; Zhang, Zhen et al. (2015). CRISPR/Cas9-mediated gene editing in human triploid zygotes. *Protein & cell* 6 (5), 363–372.
- [71] Pearson, Sue; Jia, Hepeng; Kandachi, Keiko (2004). China approves first gene therapy. *Nat Biotechnol* 22 (1), 3–4.
- [72] Ylä-Herttuala, Seppo (2012). Endgame: glybera finally recommended for approval as the first gene therapy drug in the European union. *Molecular therapy : the journal of the American Society of Gene Therapy* 20 (10), 1831–1832.
- [73] Ylä-Herttuala, Seppo (2015). Glybera's second act: the curtain rises on the high cost of therapy. *Molecular therapy : the journal of the American Society of Gene Therapy* 23 (2), 217–218.
- [74] Osborn, Mark J.; Starker, Colby G.; McElroy, Amber N.; Webber, Beau R.; Riddle, Megan J.; Xia, Lily et al. (2013). TALEN-based gene correction for epidermolysis bullosa. *Molecular therapy : the journal of the American Society of Gene Therapy* 21 (6), 1151–1159.
- [75] Ousterout, David G.; Perez-Pinera, Pablo; Thakore, Pratiksha I.; Kabadi, Ami M.; Brown, Matthew T.; Qin, Xiaoxia et al. (2013). Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. *Molecular therapy : the journal of the American Society of Gene Therapy* 21 (9), 1718–1726.
- [76] Schwank, Gerald; Koo, Bon-Kyoung; Sasselli, Valentina; Dekkers, Johanna F.; Heo, Inha; Demircan,



- Turan et al. (2013). Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell stem cell* 13 (6), 653–658.
- [77] Sebastiano, Vittorio; Maeder, Morgan L.; Angstman, James F.; Haddad, Bahareh; Khayter, Cyd; Yeo, Dana T. et al. (2011). In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem cells (Dayton, Ohio)* 29 (11), 1717–1726.
- [78] Yusa, Kosuke; Rashid, S. Tamir; Strick-Marchand, Helene; Varela, Ignacio; Liu, Pei-Qi; Paschon, David E. et al. (2011). Targeted gene correction of  $\alpha$ 1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* 478 (7369), 391–394.
- [79] Zou, Jizhong; Sweeney, Colin L.; Chou, Bin-Kuan; Choi, Uimook; Pan, Jason; Wang, Hongmei et al. (2011). Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. *Blood* 117 (21), 5561–5572.
- [80] Allers, Kristina; Hütter, Gero; Hofmann, Jörg; Loddenkemper, Christoph; Rieger, Kathrin; Thiel, Eckhard; Schneider, Thomas (2011). Evidence for the cure of HIV infection by CCR5 $\Delta$ 32/ $\Delta$ 32 stem cell transplantation. *Blood* 117 (10), 2791–2799.
- [81] Miriam Ruhenstroth (2014). Die Reparatur der Natur. In: *Die Zeit* 2014, 23.10.2014 (44).
- [82] Thierry, A.; Dujon, B. (1992). Nested chromosomal fragmentation in yeast using the meganuclease I-Sce I: a new method for physical mapping of eukaryotic genomes. *Nucleic Acids Res.* 20 (21), 5625–5631.
- [83] Silva, George; Poirot, Laurent; Galetto, Roman; Smith, Julianne; Montoya, Guillermo; Duchateau, Philippe; Paques, Frederic (2011). Meganucleases and Other Tools for Targeted Genome Engineering: Perspectives and Challenges for Gene Therapy. *CGT* 11 (1), 11–27.
- [84] Smith, Julianne; Grizot, Sylvestre; Arnould, Sylvain; Duclert, Aymeric; Epinat, Jean-Charles; Chames, Patrick et al. (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.* 34 (22), e149.
- [85] Schleifman, Erica B.; Chin, Joanna Y.; Glazer, Peter M. (2008). Triplex-mediated gene modification. *Methods in molecular biology (Clifton, N.J.)* 435, 175–190.
- [86] Gaj, Thomas; Gersbach, Charles A.; Barbas, Carlos F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology* 31 (7), 397–405.
- [87] Segal, David J.; Meckler, Joshua F. (2013). Genome engineering at the dawn of the golden age. *Annu Rev Genomics Hum Genet* 14, 135–158.
- [88] Perez-Pinera, Pablo; Ousterout, David G.; Gersbach, Charles A. (2012). Advances in targeted genome editing. *Current opinion in chemical biology* 16 (3-4), 268–277.
- [89] Urnov, Fyodor D.; Rebar, Edward J.; Holmes, Michael C.; Zhang, H. Steve; Gregory, Philip D. (2010). Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* 11 (9), 636–646.
- [90] Kim, Y. G.; Cha, J.; Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U.S.A.* 93 (3), 1156–1160.
- [91] Bitinaite, J.; Wah, D. A.; Aggarwal, A. K.; Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci. U.S.A.* 95 (18), 10570–10575.
- [92] Ramirez, Cherie L.; Foley, Jonathan E.; Wright, David A.; Müller-Lerch, Felix; Rahman, Shamim H.; Cornu, Tatjana I. et al. (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Meth* 5 (5), 374–375.
- [93] Cornu, Tatjana I.; Thibodeau-Beganny, Stacey; Guhl, Eva; Alwin, Stephen; Eichinger, Magdalena; Joung, J. Keith et al. (2008). DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Molecular therapy : the journal of the American Society of Gene Therapy* 16 (2), 352–358.
- [94] Maeder, Morgan L.; Thibodeau-Beganny, Stacey; Osiaik, Anna; Wright, David A.; Anthony, Reshma M.; Eichinger, Magdalena et al. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular cell* 31 (2), 294–301.
- [95] Sander, Jeffrey D.; Dahlborg, Elizabeth J.; Goodwin, Mathew J.; Cade, Lindsay; Zhang, Feng; Cifuentes, Daniel et al. (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Meth* 8 (1), 67–69.
- [96] Bhakta, Mital S.; Henry, Isabelle M.; Ousterout, David G.; Das, Kumitaa Theva; Lockwood, Sarah H.; Meckler, Joshua F. et al. (2013). Highly active zinc-finger nucleases by extended modular assembly. *Genome Res.* 23 (3), 530–538.
- [97] Bae, Kwang-Hee; Kwon, Young Do; Shin, Hyun-Chul; Hwang, Moon-Sun; Ryu, Eun-Hyun; Park, Kyung-Soon et al. (2003). Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* 21 (3), 275–280.
- [98] Segal, D. J.; Dreier, B.; Beerli, R. R.; Barbas, C. F. (1999). Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl. Acad. Sci. U.S.A.* 96 (6), 2758–2763.
- [99] Agrios CN (2005). Plant Pathology. San Diego: Acad. Press.
- [100] Büttner, Daniela; Bonas, Ulla (2002). Port of entry--the type III secretion translocon. *Trends in microbiology* 10 (4), 186–192.

- [101] Kay, Sabine; Bonas, Ulla (2009). How Xanthomonas type III effectors manipulate the host plant. *Current opinion in microbiology* 12 (1), 37–43.
- [102] Bonas, U.; Stall, R. E.; Staskawicz, B. (1989). Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Molecular & general genetics : MGG* 218 (1), 127–136.
- [103] Hopkins, C. M.; White, F. F.; Choi, S. H.; Guo, A.; Leach, J. E. (1992). Identification of a family of avirulence genes from *Xanthomonas oryzae* pv. *oryzae*. *Molecular plant-microbe interactions : MPMI* 5 (6), 451–459.
- [104] Kay, Sabine; Hahn, Simone; Marois, Eric; Hause, Gerd; Bonas, Ulla (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318 (5850), 648–651.
- [105] Römer, Patrick; Hahn, Simone; Jordan, Tina; Strauss, Tina; Bonas, Ulla; Lahaye, Thomas (2007). Plant pathogen recognition mediated by promoter activation of the pepper *Bs3* resistance gene. *Science* 318 (5850), 645–648.
- [106] Moscou, Matthew J.; Bogdanove, Adam J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326 (5959), 1501.
- [107] Boch, Jens; Scholze, Heidi; Schornack, Sebastian; Landgraf, Angelika; Hahn, Simone; Kay, Sabine et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326 (5959), 1509–1512.
- [108] Geißler, René; Scholze, Heidi; Hahn, Simone; Streubel, Jana; Bonas, Ulla; Behrens, Sven-Erik et al. (2011). Transcriptional Activators of Human Genes with Programmable DNA-Specificity. *PLoS ONE* 6 (5), e19509.
- [109] Garg, Abhishek; Lohmueller, Jason J.; Silver, Pamela A.; Armel, Thomas Z. (2012). Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res.* 40 (15), 7584–7595.
- [110] Miller, Jeffrey C.; Tan, Siyuan; Qiao, Guijuan; Barlow, Kyle A.; Wang, Jianbin; Xia, Danny F. et al. (2010). A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29 (2), 143–148.
- [111] Cermak, Tomas; Doyle, Erin L.; Christian, Michelle; Wang, Li; Zhang, Yong; Schmidt, Clarice et al. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39 (12), e82.
- [112] Reyon, Deepak; Tsai, Shengdar Q.; Khayter, Cyd; Foden, Jennifer A.; Sander, Jeffrey D.; Joung, J. Keith (2012). FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* 30 (5), 460–465.
- [113] Schmid-Burgk, Jonathan L.; Schmidt, Tobias; Kaiser, Vera; Höning, Klara; Hornung, Veit (2013). A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol* 31 (1), 76–81.
- [114] Briggs, Adrian W.; Rios, Xavier; Chari, Raj; Yang, Lohan; Zhang, Feng; Mali, Prashant; Church, George M. (2012). Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res.* 40 (15), e117.
- [115] Sun, Ning; Zhao, Huimin (2013). Transcription activator-like effector nucleases (TALENs): A highly efficient and versatile tool for genome editing. *Biotechnol. Bioeng.*, n/a.
- [116] Joung, J. Keith; Sander, Jeffrey D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* 14 (1), 49–55.
- [117] Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology* 169 (12), 5429–5433.
- [118] Nakata, A.; Amemura, M.; Makino, K. (1989). Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *Journal of bacteriology* 171 (6), 3553–3556.
- [119] Groenen, P. M.; Bunschoten, A. E.; van Soolingen, D.; van Embden, J D (1993). Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Molecular microbiology* 10 (5), 1057–1065.
- [120] Mojica, F. J.; Ferrer, C.; Juez, G.; Rodríguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Molecular microbiology* 17 (1), 85–93.
- [121] Masepohl, B.; Görlitz, K.; Böhme, H. (1996). Long tandemly repeated repetitive (LTRR) sequences in the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Biochimica et biophysica acta* 1307 (1), 26–30.
- [122] Hoe, N.; Nakashima, K.; Grigsby, D.; Pan, X.; Dou, S. J.; Naidich, S. et al. (1999). Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains. *Emerging infectious diseases* 5 (2), 254–263.
- [123] Mojica, F. J.; Díez-Villaseñor, C.; Soria, E.; Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology* 36 (1), 244–246.
- [124] Jansen, Ruud; Embden, Jan D A van; Gastra, Wim; Schouls, Leo M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* 43 (6), 1565–1575.
- [125] Mojica, Francisco J M; Díez-Villaseñor, César; García-Martínez, Jesús; Soria, Elena (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* 60 (2), 174–182.
- [126] Pourcel, C.; Salvignol, G.; Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and

- provide additional tools for evolutionary studies. *Microbiology (Reading, England)* 151 (Pt 3), 653–663.
- [127]** Bolotin, Alexander; Quinquis, Benoit; Sorokin, Alexei; Ehrlich, S. Dusko (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)* 151 (Pt 8), 2551–2561.
- [128]** Barrangou, Rodolphe; Fremaux, Christophe; Deveau, H el ene; Richards, Melissa; Boyaval, Patrick; Moineau, Sylvain et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315 (5819), 1709–1712.
- [129]** Mojica, F J M; D iez-Villase nor, C.; Garc ıa-Mart ınez, J.; Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading, England)* 155 (Pt 3), 733–740.
- [130]** Garneau, Josiane E.; Dupuis, Marie- eve; Villion, Manuela; Romero, Dennis A.; Barrangou, Rodolphe; Boyaval, Patrick et al. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468 (7320), 67–71.
- [131]** Brouns, Stan J J; Jore, Matthijs M.; Lundgren, Magnus; Westra, Edze R.; Slijkhuis, Rik J H; Snijders, Ambrosius P L et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321 (5891), 960–964.
- [132]** Sapranaukas, Rimantas; Gasiunas, Giedrius; Fremaux, Christophe; Barrangou, Rodolphe; Horvath, Philippe; Siksnys, Virginijus (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.* 39 (21), 9275–9282.
- [133]** Makarova, Kira S.; Haft, Daniel H.; Barrangou, Rodolphe; Brouns, Stan J J; Charpentier, Emmanuelle; Horvath, Philippe et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology* 9 (6), 467–477.
- [134]** Gesner, Emily M.; Schellenberg, Matthew J.; Garside, Erin L.; George, Mark M.; Macmillan, Andrew M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nature structural & molecular biology* 18 (6), 688–692.
- [135]** Haurwitz, Rachel E.; Jinek, Martin; Wiedenheft, Blake; Zhou, Kaihong; Doudna, Jennifer A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329 (5997), 1355–1358.
- [136]** Sashital, Dipali G.; Jinek, Martin; Doudna, Jennifer A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nature structural & molecular biology* 18 (6), 680–687.
- [137]** Deltcheva, Elitza; Chylinski, Krzysztof; Sharma, Cynthia M.; Gonzales, Karine; Chao, Yanjie; Pirzada, Zaid A. et al. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* (471), 602–607.
- [138]** Carte, Jason; Wang, Ruiying; Li, Hong; Terns, Rebecca M.; Terns, Michael P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* 22 (24), 3489–3496.
- [139]** Hale, Caryn R.; Zhao, Peng; Olson, Sara; Duff, Michael O.; Graveley, Brenton R.; Wells, Lance et al. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139 (5), 945–956.
- [140]** Marraffini, Luciano A.; Sontheimer, Erik J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322 (5909), 1843–1845.
- [141]** Manica, Andrea; Zebec, Ziga; Teichmann, Daniela; Schleper, Christa (2011). In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Molecular microbiology* 80 (2), 481–491.
- [142]** Marraffini, Luciano A.; Sontheimer, Erik J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463 (7280), 568–571.
- [143]** Jinek, Martin; Chylinski, Krzysztof; Fonfara, Ines; Hauer, Michael; Doudna, Jennifer A.; Charpentier, Emmanuelle (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337 (6096), 816–821.
- [144]** Mali, Prashant; Yang, Luhan; Esvelt, Kevin M.; Aach, John; Guell, Marc; DiCarlo, James E. et al. (2013). RNA-guided human genome engineering via Cas9. *Science* 339 (6121), 823–826.
- [145]** Shan, Qiwei; Wang, Yanpeng; Li, Jun; Zhang, Yi; Chen, Kunling; Liang, Zhen et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* 31 (8), 686–688.
- [146]** Hwang, Woong Y.; Fu, Yanfang; Reyon, Deepak; Maeder, Morgan L.; Kaini, Prakriti; Sander, Jeffrey D. et al. (2013). Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System. *PLoS ONE* 8 (7), e68708.
- [147]** Wang, Haoyi; Yang, Hui; Shivalila, Chikdu S.; Dawlaty, Meelad M.; Cheng, Albert W.; Zhang, Feng; Jaenisch, Rudolf (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153 (4), 910–918.
- [148]** Chen, Changchun; Fenk, Lorenz A.; Bono, Mario de (2013). Efficient genome editing in Caenorhabditis elegans by CRISPR-targeted homologous recombination. *Nucleic Acids Res.* 41 (20), e193.
- [149]** Gratz, Scott J.; Cummings, Alexander M.; Nguyen, Jennifer N.; Hamm, Danielle C.; Donohue, Laura K.; Harrison, Melissa M. et al. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* 194 (4), 1029–1035.
- [150]** Cong, Le; Ran, F. Ann; Cox, David; Lin, Shuailiang; Barretto, Robert; Habib, Naomi et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339 (6121), 819–823.

- [151] Cho, Seung Woo; Kim, Sojung; Kim, Jong Min; Kim, Jin-Soo (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31 (3), 230–232.
- [152] Heler, Robert; Samai, Poulami; Modell, Joshua W.; Weiner, Catherine; Goldberg, Gregory W.; Bikard, David; Marraffini, Luciano A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* 519 (7542), 199–202.
- [153] Cho, Seung Woo; Kim, Sojung; Kim, Yongsu; Kweon, Jiyeon; Kim, Heon Seok; Bae, Sangsu; Kim, Jin-Soo (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24 (1), 132–141.
- [154] Fu, Yanfang; Foden, Jennifer A.; Khayter, Cyd; Maeder, Morgan L.; Reyon, Deepak; Joung, J. Keith; Sander, Jeffrey D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*.
- [155] Fu, Yanfang; Sander, Jeffrey D.; Reyon, Deepak; Cascio, Vincent M.; Joung, J. Keith (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*.
- [156] Shen, Bin; Zhang, Wensheng; Zhang, Jun; Zhou, Jiankui; Wang, Jianying; Chen, Li et al. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Meth* 11 (4), 399–402.
- [157] Ran, F. Ann; Hsu, Patrick D.; Lin, Chie-Yu; Gootenberg, Jonathan S.; Konermann, Silvana; Trevino, Alexandro E. et al. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154 (6), 1380–1389.
- [158] Francis Hamilton (1822). An account of the fishes found in the river Ganges and its branches. Edinburgh: George Ramsey & Co.
- [159] Nelson, Joseph S. (2006). Fishes of the world. 4. Aufl. Hoboken, N.J.: J. Wiley.
- [160] Bhat, A. (2003). Diversity and composition of freshwater fishes in river systems of Central Western Ghats, India. *Environmental Biology of Fishes* (68), 25–38.
- [161] Talwar, P. K.; Jhingran, A. G. (1992). Inland fishes of India and adjacent countries. Rotterdam: A.A. Balkema.
- [162] Kinth, Priyamvadah; Mahesh, Gopalakrishnan; Panwar, Yatish (2013). Mapping of zebrafish research: a global outlook. *Zebrafish* 10 (4), 510–517.
- [163] Collodi, Paul (2004). A unique journal for a unique experimental model. *Zebrafish* 1 (1), 1.
- [164] Grunwald, David Jonah; Eisen, Judith S. (2002). Headwaters of the zebrafish -- emergence of a new model vertebrate. *Nat. Rev. Genet.* 3 (9), 717–724.
- [165] Streisinger, George; Walker, Charline; Dower, Nancy; Knauber, Donna; Singer, Fred (1981). Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature* 291 (5813), 293–296.
- [166] Schulte-Merker, S.; van Eeden, F. J.; Halpern, M. E.; Kimmel, C. B.; Nüsslein-Volhard, C. (1994). no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development* 120 (4), 1009–1015.
- [167] Rubinstein, Amy L. (2003). Zebrafish: from disease modeling to drug discovery. *Curr Opin Drug Discov Devel* 6 (2), 218–223.
- [168] Nüsslein-Volhard & Diever (1996). The Zebrafish Issue. 123. Aufl. (Development).
- [169] Kimmel, C. B.; Ballard, W. W.; Kimmel, S. R.; Ullmann, B.; Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Dev. Dyn.* 203 (3), 253–310.
- [170] Postlethwait, J. H.; Johnson, S. L.; Midson, C. N.; Talbot, W. S.; Gates, M.; Ballinger, E. W. et al. (1994). A genetic linkage map for the zebrafish. *Science* 264 (5159), 699–703.
- [171] Johnson, S. L.; Gates, M. A.; Johnson, M.; Talbot, W. S.; Horne, S.; Baik, K. et al. (1996). Centromere-linkage analysis and consolidation of the zebrafish genetic map. *Genetics* 142 (4), 1277–1288.
- [172] Postlethwait, J. (1997). Zebrafish genomics: From mutants to genes. *Trends in Genetics* 13 (5), 183–190.
- [173] Postlethwait, J. H.; Yan, Y. L.; Gates, M. A.; Horne, S.; Amores, A.; Brownlie, A. et al. (1998). Vertebrate genome evolution and the zebrafish gene map. *Nat. Genet.* 18 (4), 345–349.
- [174] Amsterdam, A.; Burgess, S.; Golling, G.; Chen, W.; Sun, Z.; Townsend, K. et al. (1999). A large-scale insertional mutagenesis screen in zebrafish. *Genes Dev.* 13 (20), 2713–2724.
- [175] Talbot, W. S.; Hopkins, N. (2000). Zebrafish mutations and functional analysis of the vertebrate genome. *Genes Dev.* 14 (7), 755–762.
- [176] Nasevicius, A.; Ekker, S. C. (2000). Effective targeted gene 'knockdown' in zebrafish. *Nat. Genet.* 26 (2), 216–220.
- [177] Howe, Kerstin; Clark, Matthew D.; Torroja, Carlos F.; Torrance, James; Berthelot, Camille; Muffato, Matthieu et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496 (7446), 498–503.
- [178] Dooley, K.; Zon, L. I. (2000). Zebrafish: a model system for the study of human disease. *Curr. Opin. Genet. Dev.* 10 (3), 252–256.
- [179] Shin, Jordan T.; Fishman, Mark C. (2002). From Zebrafish to human: modular medical models. *Annu Rev Genomics Hum Genet* 3, 311–340.
- [180] Kettleborough, Ross N W; Busch-Nentwich, Elisabeth M.; Harvey, Steven A.; Dooley, Christopher M.; Bruijn, Ewart de; van Eeden, Freek et al. (2013). A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature* 496 (7446), 494–497.
- [181] Rasooly, Rebekah S.; Henken, Deborah; Freeman, Nancy; Tompkins, Laurie; Badman, David; Briggs, Josephine; Hewitt, A. Tyl (2003). Genetic and genomic

- tools for zebrafish research: the NIH zebrafish initiative. *Dev. Dyn.* 228 (3), 490–496.
- [182] Huang, Peng; Xiao, An; Zhou, Mingguo; Zhu, Zuoyan; Lin, Shuo; Zhang, Bo (2011). Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol* 29 (8), 699–700.
- [183] Spence, Rowena; Gerlach, Gabriele; Lawrence, Christian; Smith, Carl (2008). The behaviour and ecology of the zebrafish, *Danio rerio*. *Biol Rev Camb Philos Soc* 83 (1), 13–34.
- [184] Gupta, Tripti; Mullins, Mary C. (2010). Dissection of Organs from the Adult Zebrafish. *JoVE* (37).
- [185] Truett, G. E.; Heeger, P.; Mynatt, R. L.; Truett, A. A.; Walker, J. A.; Warman, M. L. (2000). Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotech.* 29 (1), 52, 54.
- [186] Meeker, Nathan D.; Hutchinson, Sarah A.; Ho, Linh; Trede, Nikolaus S. (2007). Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *Biotech.* 43 (5), 610, 612, 614.
- [187] Westerfield, M. (2007). THE ZEBRAFISH BOOK. A guide for the laboratory use of zebrafish (*Danio rerio*). 5th Edition. Eugene: University of Oregon Press.
- [188] Mullis, K.; Faloona, F.; Scharf, S.; Saiki, R.; Horn, G.; Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51 Pt 1, 263–273.
- [189] Sambrook, J.; Fritsch, E. F.; Maniatis, T. (1989). Molecular cloning: Cold Spring Harbor laboratory press, New York.
- [190] Sanger, F.; Nicklen, S.; Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74 (12), 5463–5467.
- [191] Schroeder, Andreas; Mueller, Odilo; Stocker, Susanne; Salowsky, Ruediger; Leiber, Michael; Gassmann, Marcus et al. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* 7 (1), 3.
- [192] Huang, Peng; Zhu, Zuoyan; Lin, Shuo; Zhang, Bo (2012). Reverse Genetic Approaches in Zebrafish. *Journal of Genetics and Genomics* 39 (9), 421–433.
- [193] Doyle, E. L.; Booher, N. J.; Standage, D. S.; Voytas, D. F.; Brendel, V. P.; VanDyk, J. K.; Bogdanove, A. J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.* 40 (W1), W117.
- [194] Sanjana, Neville E.; Cong, Le; Zhou, Yang; Cunniff, Margaret M.; Feng, Guoping; Zhang, Feng (2012). A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* 7 (1), 171–192.
- [195] Sander, Jeffrey D.; Zaback, Peter; Joung, J. Keith; Voytas, Daniel F.; Dobbs, Drena (2007). Zinc Finger Targeter (ZiFIT): an engineered zinc finger/target site design tool. *Nucleic Acids Res.* 35 (Web Server issue), W599-605.
- [196] Sander, Jeffrey D.; Maeder, Morgan L.; Reyon, Deepak; Voytas, Daniel F.; Joung, J. Keith; Dobbs, Drena (2010). ZiFIT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res.* 38 (Web Server issue), W462-8.
- [197] The NCBI Sequence Read Archive. Online verfügbar unter <http://www.ncbi.nlm.nih.gov/Traces/sra/>.
- [198] Miller, Jason R.; Koren, Sergey; Sutton, Granger (2010). Assembly algorithms for next-generation sequencing data. *Genomics* 95 (6), 315–327.
- [199] Altschul, S. F.; Gish, W.; Miller, W.; Myers EW; Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol.* (251), 403–410.
- [200] Halpern, Marnie E.; Ho, Robert K.; Walker, Charline; Kimmel, Charles B. (1993). Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. *Cell* 75 (1), 99–111.
- [201] Maddison, D. R. and K.-S. Schulz (2007). The Tree of Life Web Project. Online verfügbar unter <http://tolweb.org>.
- [202] Delsuc, Frédéric; Brinkmann, Henner; Chourrout, Daniel; Philippe, Hervé (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439 (7079), 965–968.
- [203] Telford, Maximilian J.; Bourlat, Sarah J.; Economou, Andrew; Papillon, Daniel; Rota-Stabelli, Omar (2008). The evolution of the Ecdysozoa. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 363 (1496), 1529–1537.
- [204] Aguinaldo, A. M.; Turbeville, J. M.; Linford, L. S.; Rivera, M. C.; Garey, J. R.; Raff, R. A.; Lake, J. A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387 (6632), 489–493.
- [205] Chapman, Arthur D. (2009). Numbers of living species in Australia and the world. 2nd ed. Parkes, ACT: Australian Govt., Dept. of the Environment, Water, Heritage, and the Arts.
- [206] Novotny, Vojtech; Basset, Yves; Miller, Scott E.; Weiblen, George D.; Bremer, Birgitta; Cizek, Lukas; Drozd, Pavel (2002). Low host specificity of herbivorous insects in a tropical forest. *Nature* 416 (6883), 841–844.
- [207] Reaka-Kudla, Marjorie L. (1997). Biodiversity. Washington, DC: National Academy Pr.
- [208] Carde, Ring T.; Resh, Vincent H. (2009). Encyclopedia of insects. Previous ed.: 2002. 2nd ed. Amsterdam: Elsevier/Academic Press.
- [209] Footitt, R.; Adler, Peter H. (2009). Insect biodiversity. Science and society. Chichester, UK, Hoboken, NJ: Wiley-Blackwell.
- [210] Westheide W., Rieger G. (2013). Spezielle Zoologie. Teil!: Einzeller und Wirbellose. Unter Mitarbeit von Gerd Alberti. 3. Aufl. Berlin, Heidelberg: Springer Spektrum (Lehrbuch).

- [211] Wehner, Rüdiger; Gehring, Walter; Kühn, Alfred (2007). Zoologie. 24., vollst. überarb. Aufl. Stuttgart: Thieme.
- [212] Schäfer, Daniel (2012). STK33 - Evolutionäre Entwicklung in Vertebraten. Diplomarbeit. Johannes Gutenberg Universität, Mainz. Institut für Molekulargenetik.
- [213] Steffen Lerch (2013). Analyse des Gens für die Serin/Threonin Kinase 33 (Stk33) im Genom des Strauß (Struthio camelus). Diplomarbeit. Johannes Gutenberg Universität, Mainz. Institut für Molekulargenetik.
- [214] Hunter, Sarah; Jones, Philip; Mitchell, Alex; Apweiler, Rolf; Attwood, Teresa K.; Bateman, Alex et al. (2012). InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* 40 (Database issue), D306-12.
- [215] Jarvis, E. D.; Mirarab, S.; Aberer, A. J.; Li, B.; Houde, P.; Li, C. et al. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346 (6215), 1320–1331.
- [216] Jetz, W.; Thomas, G. H.; Joy, J. B.; Hartmann, K.; Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature* 491 (7424), 444–448.
- [217] Zhang, G.; Li, C.; Li, Q.; Li, B.; Larkin, D. M.; Lee, C. et al. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346 (6215), 1311–1320.
- [218] Schröder, Rienovita (2014). *stk33* Expression in Vögeln. Diplomarbeit. Johannes Gutenberg Universität Mainz, Mainz. Institut für Molekulargenetik.
- [219] Gabriel, Richard; Lombardo, Angelo; Arens, Anne; Miller, Jeffrey C.; Genovese, Pietro; Kaeppl, Christine et al. (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* 29 (9), 816–823.
- [220] Mussolino, Claudio; Morbitzer, Robert; Lütge, Fabienne; Dannemann, Nadine; Lahaye, Thomas; Cathomen, Toni (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* 39 (21), 9283–9293.
- [221] Pattanayak, Vikram; Ramirez, Cherie L.; Joung, J. Keith; Liu, David R. (2011). Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Meth* 8 (9), 765–770.
- [222] Pattanayak, Vikram; Lin, Steven; Guilinger, John P.; Ma, Enbo; Doudna, Jennifer A.; Liu, David R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 31 (9), 839–843.
- [223] Sahdev, Sudhir; Saini, Shalini; Tiwari, Prabhakar; Saxena, Sanjeev; Singh Saini, Kulvinder (2007). Amplification of GC-rich genes by following a combination strategy of primer design, enhancers and modified PCR cycle conditions. *Molecular and cellular probes* 21 (4), 303–307.
- [224] Hommelshaus, Carl Maximilian; Frantzeskakis, Lamprinos; Huang, Mengmeng; Ülker, Bekir (2014). PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. *Sci Rep* 4, 5052.
- [225] Sander, Jeffrey D.; Cade, Lindsay; Khayter, Cyd; Reyon, Deepak; Peterson, Randall T.; Joung, J. Keith; Yeh, Jing-Ruey J. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol* 29 (8), 697–698.
- [226] Kim, Hye Joo; Lee, Hyung Joo; Kim, Hyojin; Cho, Seung Woo; Kim, Jin-Soo (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res.* 19 (7), 1279–1288.
- [227] Hsu, Patrick D.; Scott, David A.; Weinstein, Joshua A.; Ran, F. Ann; Konermann, Silvana; Agarwala, Vineeta et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31 (9), 827–832.
- [228] Gagnon, James A.; Valen, Eivind; Thyme, Summer B.; Huang, Peng; Ahkmetova, Laila; Pauli, Andrea et al. (2014). Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs. *PLoS ONE* 9 (5), e98186.
- [229] Hruscha, Alexander; Krawitz, Peter; Rechenberg, Alexandra; Heinrich, Verena; Hecht, Jochen; Haass, Christian; Schmid, Bettina (2013). Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development* 140 (24), 4982–4987.
- [230] Kimura, Yukiko; Hisano, Yu; Kawahara, Atsuo; Higashijima, Shin-Ichi (2014). Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. *Sci Rep* 4, 6545.
- [231] Liu, Da; Wang, Zhanxiang; Xiao, An; Zhang, Yutian; Li, Wenyuan; Zu, Yao et al. (2014). Efficient gene targeting in zebrafish mediated by a zebrafish-codon-optimized cas9 and evaluation of off-targeting effect. *J Genet Genomics* 41 (1), 43–46.
- [232] Seruggia, Davide; Montoliu, Lluís (2014). The new CRISPR-Cas system: RNA-guided genome engineering to efficiently produce any desired genetic alteration in animals. *Transgenic Res.*
- [233] Mali, Prashant; Aach, John; Stranges, P. Benjamin; Esvelt, Kevin M.; Moosburner, Mark; Kosuri, Sriram et al. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* 31 (9), 833–838.
- [234] Rosen, Jonathan N.; Sweeney, Michael F.; Mably, John D. (2009). Microinjection of Zebrafish Embryos to Analyze Gene Function. *JoVE* (25).
- [235] Ablain, Julien; Durand, Ellen M.; Yang, Song; Zhou, Yi; Zon, Leonard I. (2015). A CRISPR/Cas9 Vector System for Tissue-Specific Gene Disruption in Zebrafish. *Developmental Cell.*
- [236] Fayyad, Usama M. (1996). Advances in knowledge discovery and data mining. Menlo Park, Calif.: AAAI Press; MIT Press.

- [237] Wheeler, David L.; Barrett, Tanya; Benson, Dennis A.; Bryant, Stephen H.; Canese, Kathi; Chetvernin, Vyacheslav et al. (2008). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 36 (Database issue), D13–21.
- [238] Vienne, Alexandre (2006). Metaphylogeny of 82 gene families sheds a new light on chordate evolution. *Int. J. Biol. Sci.*, 32–37.
- [239] Berná, Luisa; Alvarez-Valin, Fernando (2014). Evolutionary genomics of fast evolving tunicates. *Genome biology and evolution* 6 (7), 1724–1738.
- [240] Dehal, Paramvir; Satou, Yutaka; Campbell, Robert K.; Chapman, Jarrod; Degnan, Bernard; Tomaso, Anthony de et al. (2002). The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298 (5601), 2157–2167.
- [241] Holland, Linda Z.; Gibson-Brown, Jeremy J. (2003). The *Ciona intestinalis* genome: when the constraints are off. *BioEssays: news and reviews in molecular, cellular and developmental biology* 25 (6), 529–532.
- [242] Treangen, Todd J.; Salzberg, Steven L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13 (1), 36–46.
- [243] Alkan, Can; Coe, Bradley P.; Eichler, Evan E. (2011). Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* 12 (5), 363–376.
- [244] English, Adam C.; Richards, Stephen; Han, Yi; Wang, Min; Vee, Vanesa; Qu, Jiaxin et al. (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS ONE* 7 (11), e47768.
- [245] Piro, Vitor C.; Faoro, Helisson; Weiss, Vinicius A.; Steffens, Maria B R; Pedrosa, Fabio O.; Souza, Emanuel M.; Raittz, Roberto T. (2014). FGAP: an automated gap closing tool. *BMC research notes* 7, 371.
- [246] Levene, M. J.; Korlach, J.; Turner, S. W.; Foquet, M.; Craighead, H. G.; Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299 (5607), 682–686.
- [247] Eid, John; Fehr, Adrian; Gray, Jeremy; Luong, Khai; Lyle, John; Otto, Geoff et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323 (5910), 133–138.
- [248] Hufton, Andrew L.; Groth, Detlef; Vingron, Martin; Lehrach, Hans; Poustka, Albert J.; Panopoulou, Georgia (2008). Early vertebrate whole genome duplications were predated by a period of intense genome rearrangement. *Genome Res.* 18 (10), 1582–1591.
- [249] Benton, Michael J.; Ayala, Francisco J. (2003). Dating the tree of life. *Science* 300 (5626), 1698–1700.
- [250] Blair, Jaime E.; Hedges, S. Blair (2005). Molecular clocks do not support the Cambrian explosion. *Molecular biology and evolution* 22 (3), 387–390.
- [251] Hedges, S. Blair; Blair, Jaime E.; Venturi, Maria L.; Shoe, Jason L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC evolutionary biology* 4, 2.
- [252] Peterson, Kevin J.; Lyons, Jessica B.; Nowak, Kristin S.; Takacs, Carter M.; Wargo, Matthew J.; McPeck, Mark A. (2004). Estimating metazoan divergence times with a molecular clock. *Proc. Natl. Acad. Sci. U.S.A.* 101 (17), 6536–6541.
- [253] Chen, J. Y.; Vannier, J.; Huang, D. Y. (2001). The origin of crustaceans: new evidence from the Early Cambrian of China. *Proceedings. Biological sciences / The Royal Society* 268 (1482), 2181–2187.
- [254] Penney, David; Selden, Paul (2011). Fossil spiders. The evolutionary history of a mega-diverse order. Manchester: Siri Scientific Press (Monograph series, v. 1).
- [255] Misof, Bernhard; Liu, Shanlin; Meusemann, Karen; Peters, Ralph S.; Donath, Alexander; Mayer, Christoph et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346 (6210), 763–767.
- [256] Peters, Ralph S.; Meyer, Benjamin; Krogmann, Lars; Borner, Janus; Meusemann, Karen; Schütte, Kai et al. (2011). The taming of an impossible child: a standardized all-in approach to the phylogeny of Hymenoptera using public database sequences. *BMC biology* 9, 55.
- [257] Wang, D. Y.; Kumar, S.; Hedges, S. B. (1999). Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proceedings. Biological sciences / The Royal Society* 266 (1415), 163–171.
- [258] Frank Gill, David Donsker (2014). IOC World Bird List. Online verfügbar unter <http://www.worldbirdnames.org/>.
- [259] Feduccia, Alan (2003). ‘Big bang’ for tertiary birds? *Trends in Ecology & Evolution* 18 (4), 172–176.
- [260] Lautwein, Tobias; Lerch, Steffen; Schäfer, Daniel; Schmidt, Erwin R. (2015). The serine/threonine kinase 33 is present and expressed in palaeognath birds but has become a unitary pseudogene in neognaths about 100 million years ago. *BMC genomics* 16, 543.
- [261] Shemesh, Ronen; Novik, Amit; Edelheit, Sarit; Sorek, Rotem (2006). Genomic fossils as a snapshot of the human transcriptome. *Proc. Natl. Acad. Sci. U.S.A.* 103 (5), 1364–1369.
- [262] Petrov, D. A.; Lozovskaya, E. R.; Hartl, D. L. (1996). High intrinsic rate of DNA loss in *Drosophila*. *Nature* 384 (6607), 346–349.
- [263] Zhang, Guojie; Rahbek, Carsten; Graves, Gary R.; Lei, Fumin; Jarvis, Erich D.; Gilbert, M Thomas P (2015). Genomics: Bird sequencing project takes off. *Nature* 522 (7554), 34.
- [264] Koskiniemi, Sanna; Sun, Song; Berg, Otto G.; Andersson, Dan I. (2012). Selection-driven gene loss in bacteria. *PLoS Genet* 8 (6), e1002787.

- [265] Krylov, Dmitri M.; Wolf, Yuri I.; Rogozin, Igor B.; Koonin, Eugene V. (2003). Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* 13 (10), 2229–2235.
- [266] Wolf, Yuri I.; Makarova, Kira S.; Yutin, Natalya; Koonin, Eugene V. (2012). Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biology direct* 7, 46.
- [267] Lovell, Peter V.; Wirthlin, Morgan; Wilhelm, Larry; Minx, Patrick; Lazar, Nathan H.; Carbone, Lucia et al. (2014). Conserved syntenic clusters of protein coding genes are missing in birds. *Genome Biol.* 15 (12), 117.
- [268] Hughes, Austin L.; Friedman, Robert (2008). Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes. *Molecular biology and evolution* 25 (12), 2681–2688.
- [269] Newman, Stuart A.; Mezentseva, Nadezhda V.; Badyaev, Alexander V. (2013). Gene loss, thermogenesis, and the origin of birds. *Annals of the New York Academy of Sciences* 1289, 36–47.
- [270] mGroen, Jessica N.; Capraro, David; Morris, Kevin V. (2014). The emerging role of pseudogene expressed non-coding RNAs in cellular functions. *Int. J. Biochem. Cell Biol.* 54C, 350–355.
- [271] Zhang, Zhengdong D.; Frankish, Adam; Hunt, Toby; Harrow, Jennifer; Gerstein, Mark (2010). Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biol.* 11 (3), R26.
- [272] Levine, M. (1986). New concepts in the biology and biochemistry of ascorbic acid. *N. Engl. J. Med.* 314 (14), 892–902.
- [273] Lachapelle, Marc Y.; Drouin, Guy (2011). Inactivation dates of the human and guinea pig vitamin C genes. *Genetica* 139 (2), 199–207.
- [274] Nishikimi, M.; Kawai, T.; Yagi, K. (1992). Guinea pigs possess a highly mutated gene for L-gulonogamma-lactone oxidase, the key enzyme for L-ascorbic acid biosynthesis missing in this species. *The Journal of biological chemistry* 267 (30), 21967–21972.
- [275] Birney, E. C.; Jenness, R.; Ayaz, K. M. (1976). Inability of bats to synthesise L-ascorbic acid. *Nature* 260 (5552), 626–628.
- [276] Milton, K.; Jenness, R. (1987). Ascorbic acid content of neotropical plant parts available to wild monkeys and bats. *Experientia* 43 (3), 339–342.
- [277] Drouin, Guy; Godin, Jean-Rémi; Pagé, Benoît (2011). The genetics of vitamin C loss in vertebrates. *Current genomics* 12 (5), 371–378.
- [278] Olson, M. V. (1999). When less is more: gene loss as an engine of evolutionary change. *American journal of human genetics* 64 (1), 18–23.
- [279] Olson, Maynard V.; Varki, Ajit (2003). Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat. Rev. Genet.* 4 (1), 20–28.
- [280] Galili, U.; Swanson, K. (1991). Gene sequences suggest inactivation of alpha-1,3-galactosyltransferase in catarrhines after the divergence of apes from monkeys. *Proc. Natl. Acad. Sci. U.S.A.* 88 (16), 7401–7404.
- [281] Chou, Hsun-Hua; Hayakawa, Toshiyuki; Diaz, Sandra; Krings, Matthias; Indriati, ETTY; Leakey, Meave et al. (2002). Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl. Acad. Sci. U.S.A.* 99 (18), 11736–11741.
- [282] Stedman, Hansell H.; Kozyak, Benjamin W.; Nelson, Anthony; Thesier, Danielle M.; Su, Leonard T.; Low, David W. et al. (2004). Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature* 428 (6981), 415–418.
- [283] Carter, Hannah; Samayoa, Josue; Hruban, Ralph H.; Karchin, Rachel (2010). Prioritization of driver mutations in pancreatic cancer using cancer-specific high-throughput annotation of somatic mutations (CHASM). *Cancer Biol. Ther.* 10 (6), 582–587.
- [284] Galperin, M. Y.; Walker, D. R.; Koonin, E. V. (1998). Analogous enzymes: independent inventions in enzyme evolution. *Genome Res.* 8 (8), 779–790.
- [285] Kafri, Ran; Levy, Melissa; Pilpel, Yitzhak (2006). The regulatory utilization of genetic redundancy through responsive backup circuits. *Proc. Natl. Acad. Sci. U.S.A.* 103 (31), 11653–11658.
- [286] Kafri, Ran; Springer, Michael; Pilpel, Yitzhak (2009). Genetic redundancy: new tricks for old genes. *Cell* 136 (3), 389–392.
- [287] Carracedo, Sergio; Sacher, Frank; Brandes, Gudrun; Braun, Ursula; Leitges, Michael (2014). Redundant role of protein kinase C delta and epsilon during mouse embryonic development. *PLoS ONE* 9 (8), e103686.
- [288] Enns, Linda C.; Kanaoka, Masahiro M.; Torii, Keiko U.; Comai, Luca; Okada, Kiyotaka; Cleland, Robert E. (2005). Two callose synthases, GSL1 and GSL5, play an essential and redundant role in plant and pollen development and in fertility. *Plant molecular biology* 58 (3), 333–349.
- [289] Pearce, Andrew C.; Senis, Yotis A.; Billadeau, Daniel D.; Turner, Martin; Watson, Steve P.; Vigorito, Elena (2004). Vav1 and vav3 have critical but redundant roles in mediating platelet activation by collagen. *The Journal of biological chemistry* 279 (52), 53955–53962.
- [290] Nowak, M. A.; Boerlijst, M. C.; Cooke, J.; Smith, J. M. (1997). Evolution of genetic redundancy. *Nature* 388 (6638), 167–171.
- [291] Koonin, E. V.; Mushegian, A. R.; Bork, P. (1996). Non-orthologous gene displacement. *Trends in genetics : TIG* 12 (9), 334–336.
- [292] Marsh, J. J.; Lebherz, H. G. (1992). Fructose-bisphosphate aldolases: an evolutionary history. *Trends in biochemical sciences* 17 (3), 110–113.



- [293] Rutter, W. J.A (1964). Evolution of aldolase. *Federation proceedings* 23, 1248–1257.
- [294] Perham, R. N. (1990). The fructose-1,6-bisphosphate aldolases: same reaction, different enzymes. *Biochemical Society transactions* 18 (2), 185–187.
- [295] Houten, S. M.; Waterham, H. R. (2001). Nonorthologous gene displacement of phosphomevalonate kinase. *Molecular genetics and metabolism* 72 (3), 273–276.
- [296] Yamaguchi, Tomoya; Goto, Hidemasa; Yokoyama, Tomoya; Silljé, Herman; Hanisch, Anja; Uldschmid, Andreas et al. (2005). Phosphorylation by Cdk1 induces Plk1-mediated vimentin phosphorylation during mitosis. *J. Cell Biol.* 171 (3), 431–436.
- [297] Chou, Y. H.; Opal, P.; Quinlan, R. A.; Goldman, R. D. (1996). The relative roles of specific N- and C-terminal phosphorylation sites in the disassembly of intermediate filament in mitotic BHK-21 cells. *J. Cell. Sci.* 109 (Pt 4), 817–826.
- [298] Goto, Hidemasa; Yasui, Yoshihiro; Kawajiri, Aie; Nigg, Erich A.; Terada, Yasuhiko; Tatsuka, Masaaki et al. (2003). Aurora-B regulates the cleavage furrow-specific vimentin phosphorylation in the cytokinetic process. *The Journal of biological chemistry* 278 (10), 8526–8530.
- [299] Chou, Ying-Hao; Bischoff, James R.; Beach, David; Goldman, Robert D. (1990). Intermediate filament reorganization during mitosis is mediated by p34cdc2 phosphorylation of vimentin. *Cell* 62 (6), 1063–1071.
- [300] Tsujimura, K.; Ogawara, M.; Takeuchi, Y.; Imajoh-Ohmi, S.; Ha, M. H.; Inagaki, M. (1994). Visualization and function of vimentin phosphorylation by cdc2 kinase during mitosis. *The Journal of biological chemistry* 269 (49), 31097–31106.
- [301] Chou, Y. H.; Ngai, K. L.; Goldman, R. (1991). The regulation of intermediate filament reorganization in mitosis. p34cdc2 phosphorylates vimentin at a unique N-terminal site. *The Journal of biological chemistry* 266 (12), 7325–7328.
- [302] Tang, Dale D.; Bai, Ying; Gunst, Susan J. (2005). Silencing of p21-activated kinase attenuates vimentin phosphorylation on Ser-56 and reorientation of the vimentin network during stimulation of smooth muscle cells by 5-hydroxytryptamine. *The Biochemical journal* 388 (Pt 3), 773–783.
- [303] Wang, Ruping; Li, Qing-Fen; Anfinogenova, Yana; Tang, Dale D. (2007). Dissociation of Crk-associated substrate from the vimentin network is regulated by p21-activated kinase on ACh activation of airway smooth muscle. *American journal of physiology. Lung cellular and molecular physiology* 292 (1), L240-8.
- [304] Stefanovic, Sandra; Windsor, Miriam; Nagata, Koh-Ici; Inagaki, Masaki; Wileman, Thomas (2005). Vimentin rearrangement during African swine fever virus infection involves retrograde transport along microtubules and phosphorylation of vimentin by calcium calmodulin kinase II. *Journal of virology* 79 (18), 11766–11775.
- [305] Oguri, Takashi; Inoko, Akihito; Shima, Hiroshi; Izawa, Ichiro; Arimura, Nariko; Yamaguchi, Tomoya et al. (2006). Vimentin-Ser82 as a memory phosphorylation site in astrocytes. *Genes Cells* 11 (5), 531–540.
- [306] Cheng, T. J.; Lai, Y. K. (1998). Identification of mitogen-activated protein kinase-activated protein kinase-2 as a vimentin kinase activated by okadaic acid in 9L rat brain tumor cells. *Journal of cellular biochemistry* 71 (2), 169–181.
- [307] Cheng, Ting-Jen; Tseng, Yu-Fang; Chang, Whei-Meih; Chang, Margaret Dah-Tsyr; Lai, Yiu-Kay (2003). Retaining of the assembly capability of vimentin phosphorylated by mitogen-activated protein kinase-activated protein kinase-2. *Journal of cellular biochemistry* 89 (3), 589–602.
- [308] Ogawara, M.; Inagaki, N.; Tsujimura, K.; Takai, Y.; Sekimata, M.; Ha, M. H. et al. (1995). Differential targeting of protein kinase C and CaM kinase II signalings to vimentin. *J. Cell Biol.* 131 (4), 1055–1066.
- [309] Takai, Y.; Ogawara, M.; Tomono, Y.; Moritoh, C.; Imajoh-Ohmi, S.; Tsutsumi, O. et al. (1996). Mitosis-specific phosphorylation of vimentin by protein kinase C coupled with reorganization of intracellular membranes. *J. Cell Biol.* 133 (1), 141–149.
- [310] Goto, H. (1998). Phosphorylation of Vimentin by Rho-associated Kinase at a Unique Amino-terminal Site That Is Specifically Phosphorylated during Cytokinesis. *Journal of Biological Chemistry* 273 (19), 11728–11736.
- [311] Eriksson, John E.; He, Tao; Trejo-Skalli, Amy V.; Härmälä-Braskén, Ann-Sofi; Hellman, Jukka; Chou, Ying-Hao; Goldman, Robert D. (2004). Specific in vivo phosphorylation sites determine the assembly dynamics of vimentin intermediate filaments. *J. Cell. Sci.* 117 (Pt 6), 919–932.
- [312] Herrmann, H.; Aebi, U. (1998). Structure, assembly, and dynamics of intermediate filaments. *Subcellular biochemistry* 31, 319–362.
- [313] Kochin, Vitaly; Imanishi, Susumu Y.; Eriksson, John E. (2006). Fast track to a phosphoprotein sketch - MALDI-TOF characterization of TLC-based tryptic phosphopeptide maps at femtomolar detection sensitivity. *Proteomics* 6 (21), 5676–5682.
- [314] Izawa, Ichiro; Inagaki, Masaki (2006). Regulatory mechanisms and functions of intermediate filaments: a study using site- and phosphorylation state-specific antibodies. *Cancer science* 97 (3), 167–174.

# Anhang

## A1 Vektorkarten

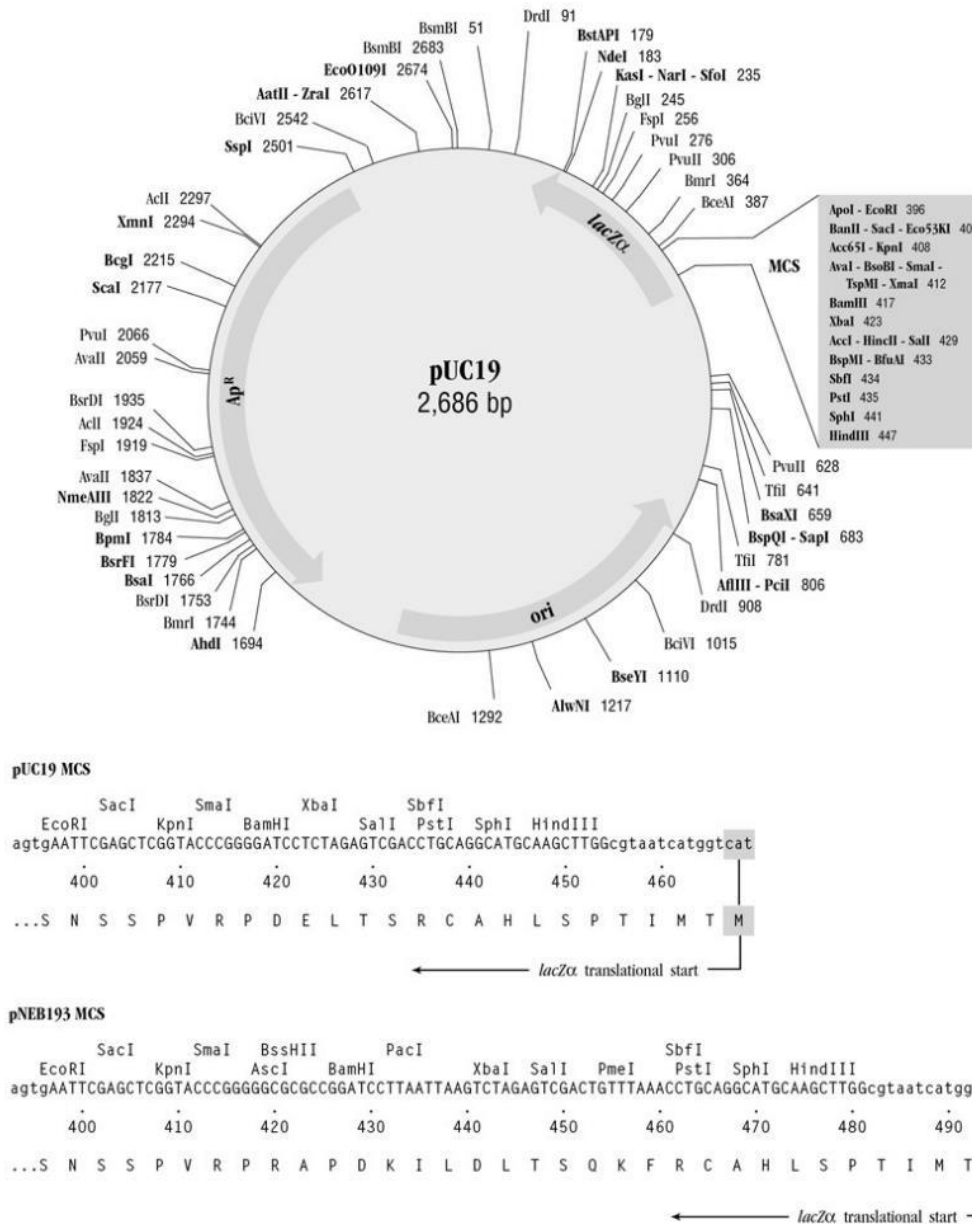


Abbildung A 1 | Übersicht über die verwendeten Vektoren. Dargestellt ist die Vektorkarte des pUC19 (NEB, Ipswich, USA).

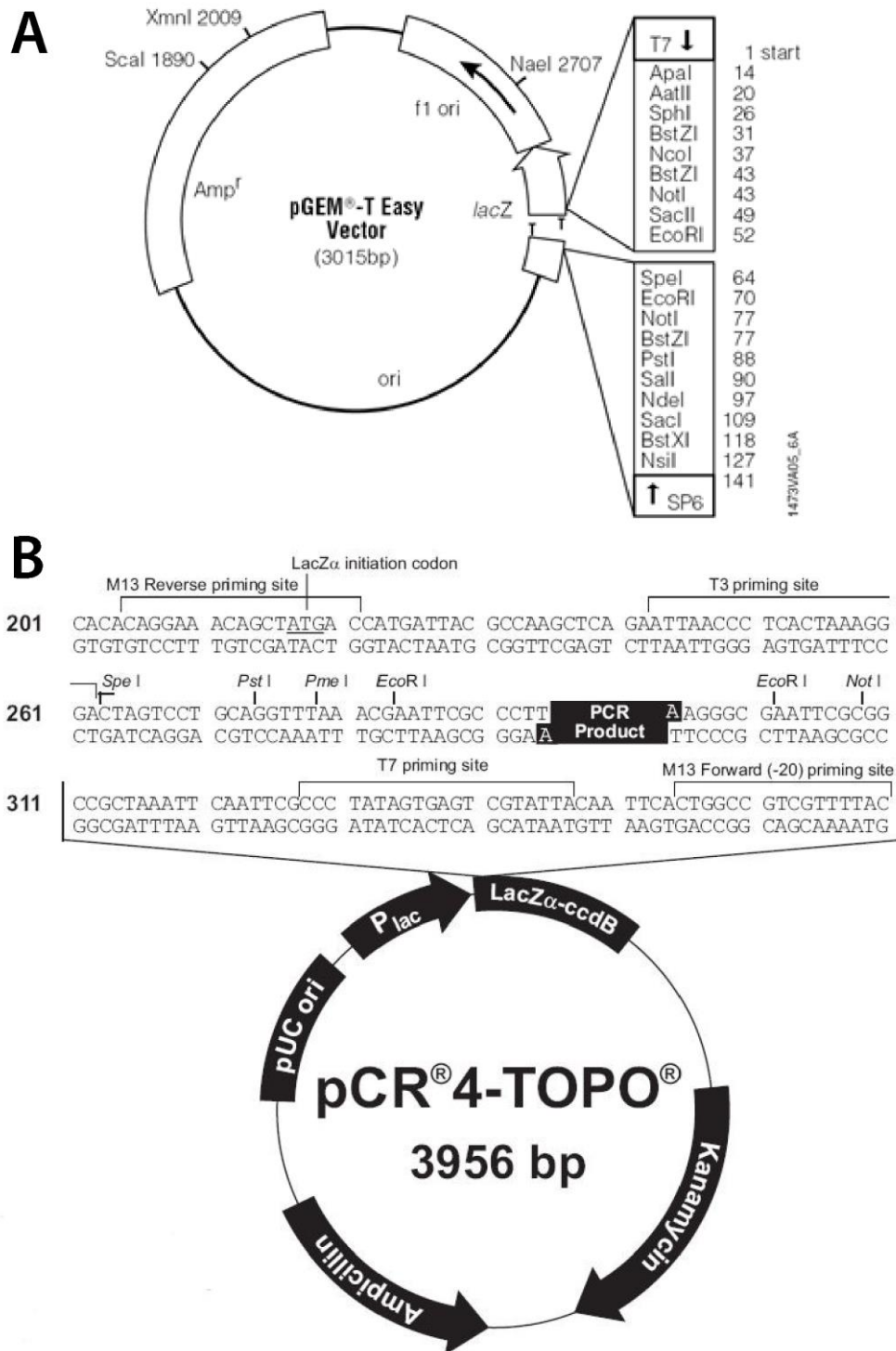
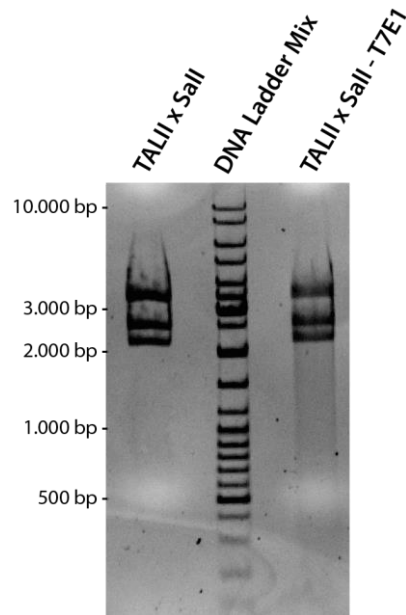
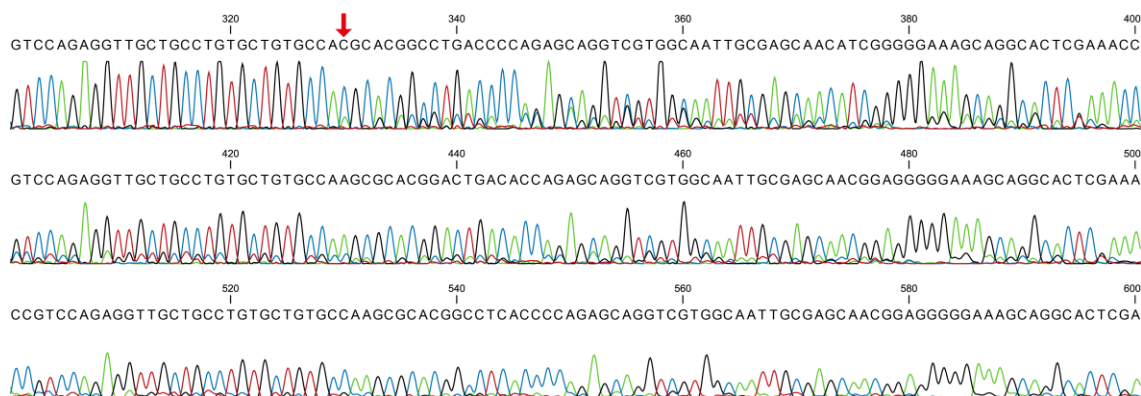


Abbildung A 2 | Übersicht über die verwendeten Vektoren. Dargestellt sind (A) die Vektorkarte des pGEM® T-Easy (Promega, Madison, USA) und (B) die Vektorkarte des pCR®4-TOPO® der Fa. Invitrogen (Carlsbad, USA).

## A2 TALEN-vermittelter Knock-out von *stk33* im Zebrafisch

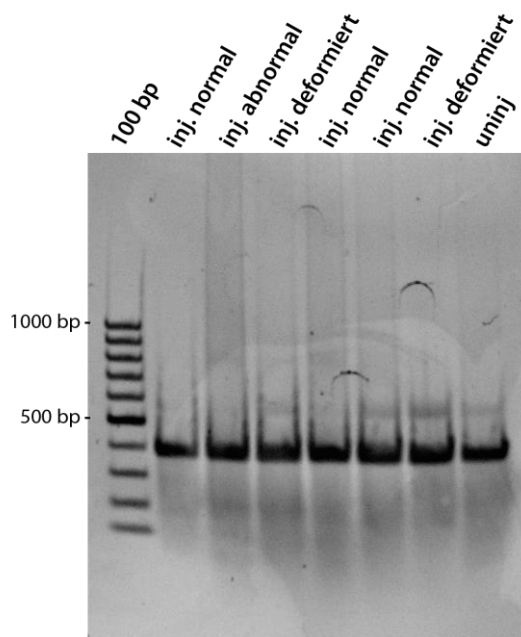


**Abbildung A 3 | Elektrophoretische Analyse des T7E1-Assay des TALII-Vektors.** Zur Überprüfung von TALII auf das Vorhandensein einer eventuellen Deletion in einem Teil der Plasmide wurde zunächst ein Verdau mit *Sall* und anschließend ein T7E1-Assay durchgeführt. Die elektrophoretische Analyse der hieraus entstandenen Fragmente zeigt jedoch lediglich die durch den *Sall*-Restriktionsverdau entstandenen Fragmente und keines der im Falle einer tatsächlich vorliegenden Deletion zu erwartenden Spaltprodukte.

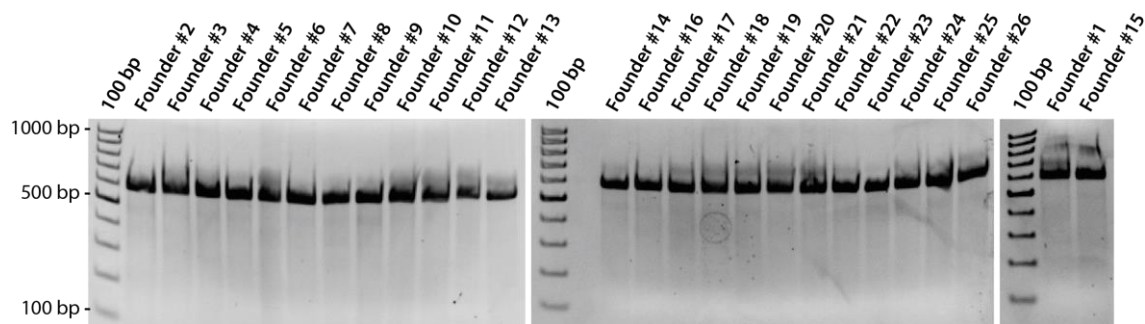


**Abbildung A 4 | Chromatogramm der Sequenzierung der assemblierten TALEN-Vektoren.** Bei der Sequenzierung vollständiger Klone zwecks Überprüfung der korrekten Reihenfolge der TALEN-Monomere kam es ab bestimmten Positionen (roter Pfeil) zu überlagernden Signalen, welche sich jedoch nach ca. 150 Basen wieder normalisierten.

## A3 CRISPR/Cas-vermittelter Knock-out von *stk33*

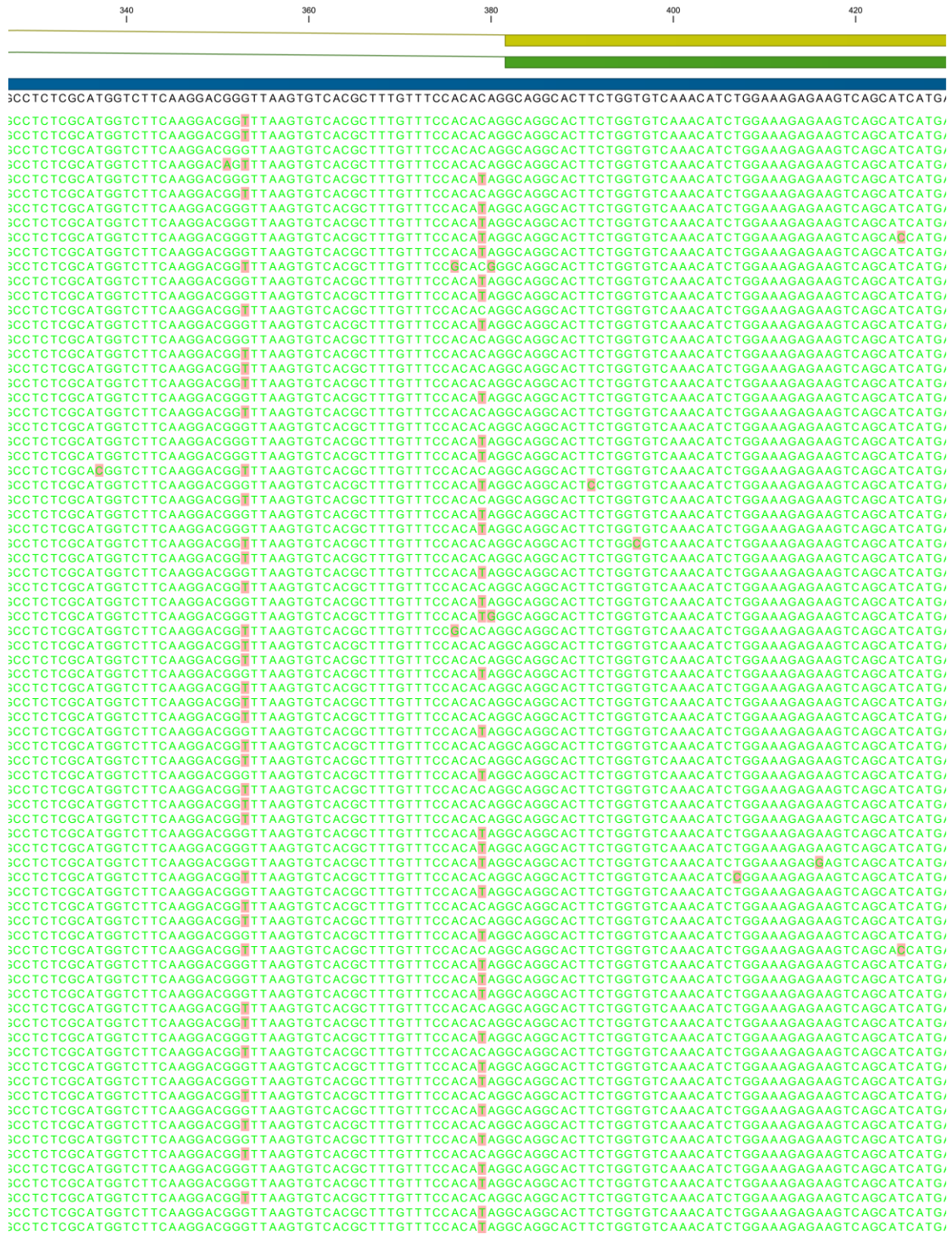


**Abbildung A 5 | Gelelektrophoretische Analyse des T7EI-Assay der *stk33*-I „off-target“ Zielsequenz.** Hierbei ist neben der Bande des ursprünglichen PCR-Produktes von 382 bp keines der zu erwartenden Spaltprodukte von ca. 170 bp und 210 bp zu erkennen.



**Abbildung A 6 | Elektrophoretische Analyse des T7EI-Assay der adulten  $F_0$ -Generation.** Bei keinem der 26 für die Auskreuzung verwendeten adulten Zebrafischen konnten die im Falle einer Mutation zu erwartenden Spaltprodukte nachgewiesen werden.





**Abbildung A 7 | Analyse des für die Genotypisierung verwendeten Bereiches hinsichtlich möglicher Varianz.** Bei der gelelektrophoretischen Analyse der injizierten Embryonen zeigten sich unerwartete Spaltprodukte von 370 bp und 210 bp in allen untersuchten Proben sowie in der uninjizierten Kontrolle. Bei der Analyse der beiden möglichen Schnittstellen mittels eines Sequenzalignments zeigten sich bei der im Intron gelegenen Schnittstellen zwei heterozygot vorliegende SNPs. Genannotationen sind in Blau, mRNA-Annotationen in Grün und kodierende Bereich in Gelb dargestellt. Fehlpaarungen sind in Rot hervorgehoben.

## A4 *Stk33* in der Klasse Aves

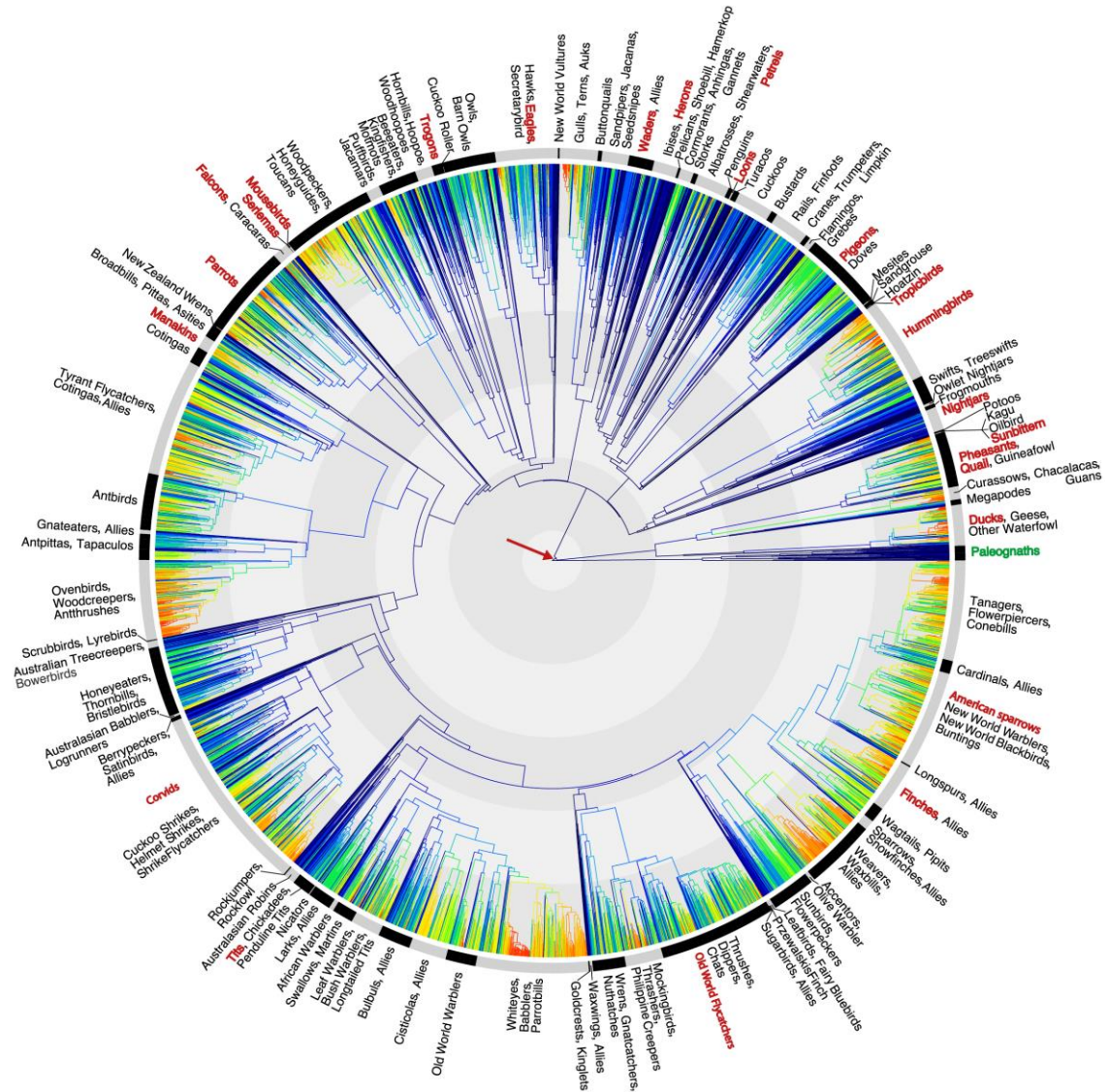
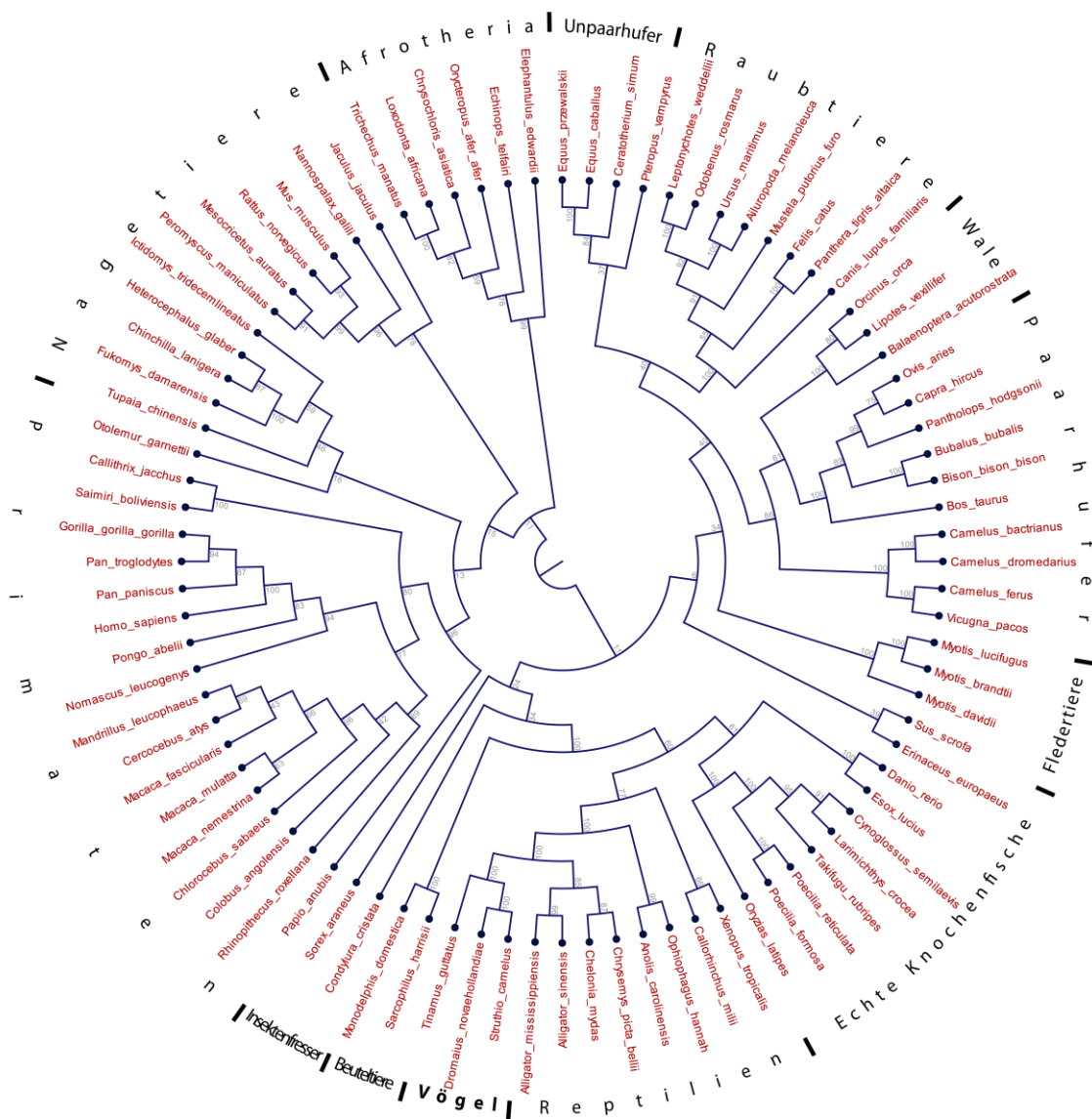


Abbildung A 8 | Phylogenetischer Stammbaum der Klasse Aves mit Kennzeichnung untersuchter Familien. Familien welche über ein funktionales *stk33* verfügen sind in Grün dargestellt, Familien mit nicht-funktionalen Überresten von *stk33* sind in Rot dargestellt. Der Verlust von *stk33* an der Verzweigung zwischen Ur- und Neukiefernvögeln ist durch einen roten Pfeil markiert. (Abbildung übernommen und verändert von [www.birdtree.org](http://www.birdtree.org) mit Zustimmung der Autoren und des ursprünglichen Urhebers, Gavin Thomas. Die Abbildung basiert auf Daten aus: Jetz, W; Thomas, GH; Joy, JB; Hartmann, K; Mooers, AO: *The global diversity of birds in space and time*. Nature 2012, 7424:444–448)



**Abbildung A 9 | Alignment identifizierter *stk33* Exons aus den verbleibenden Vogelspezies gegen die *stk33* mRNA Sequenz des Vogelstraußes.** Die mRNA Sequenz des Vogelstraußes ist als grüner Pfeil, der offene Leserahmen als gelber Pfeil dargestellt. Grüne Linien kennzeichnen kartierte Exons, rote Linien kennzeichnen invertierte Exons. Kleine Indels (<5 Bp) sind durch blaue Dreiecke, größere Indels (>5 Bp) durch blaue Rechtecke gezeigt. Stopkodons sind durch rote Sterne und Start Kodons durch den grünen Großbuchstaben „M“ angedeutet. Schwarze Linien stellen Konsensussequenzen dar. Die Position von Exon 3 ist durch eine violett gestrichelte Linie hervorgehoben.





**Abbildung A 10 | STK33 Phylogenie.** Alignment aller verifizierten STK33 Proteine aus dem Unterstamm der Wirbeltiere (*Vertebrata*) inklusive der in der Klasse der Vögel (*Aves*) gefundenen STK33-Proteine und die daraus resultierende Maximum-Likelihood-Phylogenie (Neighbor-joining), dargestellt als radiäres Kladogramm. Bootstrap Werte sind neben den entsprechenden Knoten dargestellt.

## A5 Übersicht über untersuchte Arten aus dem Reich

### *Animalia*

**Tabelle A 1 | Detaillierte Übersicht über untersuchte Arten innerhalb der *Insecta*.** Dargestellt sind die je Taxon analysierten Arten sowie die dazu verwendeten genomischen oder transkriptomischen Datensätze. Ein positiver Fund auf genomischer oder transkriptomischer Ebene ist durch ein (+) in der entsprechenden Spalte gekennzeichnet. War kein Hinweis auf *stk33* in den analysierten Daten vorhanden, so ist dies durch ein (-) gekennzeichnet.

Taxon (Klasse/Ordnung)	untersuchte Art	genomisch	transkriptomisch	gen.	trans.
Zygentoma	<i>Atelura formicaria</i>	x	GAYJ00000000		-
	<i>Tricholepidion gertschi</i>	x	GASO00000000		+
	<i>Thermobia domestica</i>	x	GASN00000000		+
Archaeognatha	<i>Meinertellus cundinamarcensis</i>	x	GAUG00000000		+
	<i>Machilis hrabei</i>	x	GAUM00000000		-
Ephemeroptera	<i>Ephemera danica</i>	PRJNA171755	GAUK00000000	-	-
	<i>Isonychia bicolor</i>	x	GAXA00000000		-
	<i>Baetis</i> sp.	x	GATU00000000		-
	<i>Eurylophella</i> sp.	x	GAZG00000000		-
Odonata	<i>Ladona fulva</i>	PRJNA194433	x	-	
	<i>Calopteryx splendens</i>	x	GAYM00000000		-
	<i>Epiophlebia superstes</i>	x	GAVW00000000		-
	<i>Cordulegaster boltonii</i>	x	GAYO00000000		-
	<i>Enallagma hageni</i>	x	GAEQ00000000		-
Plecoptera	<i>Leuctra</i> sp.	x	GAUF00000000		-
	<i>Perla marginata</i>	x	GATV00000000		-
	<i>Cosmioperla kuna</i>	x	GAYL00000000		-
Embioptera	<i>Nemurella pictetii</i>	x	GAAV00000000		-
	<i>Haploembia palaui</i>	x	GAZA00000000		-
Phasmida	<i>Aposthonia japonica</i>	x	GAWU00000000		-
	<i>Timema cristinae</i>	x	GAVX00000000		-
	<i>Peruphasma schultei</i>	x	GAWJ00000000		-
	<i>Aretaon asperrimus</i>	x	GAZQ00000000		-
	<i>Extatosoma tiaratum</i>	x	GAWG00000000		-
Orthoptera	<i>Sipyloidea sipylos</i>	x	GAWF00000000		-
	<i>Ramulus artemis</i>	x	GAWE00000000		-
	<i>Medauroidea extradentata</i>	x	GAWD00000000		-
	<i>Locusta migratoria</i>	PRJNA185471	GCGJ00000000 GBDZ00000000	+	-
	<i>Stenobothrus lineatus</i>	X	GAUZ00000000		+
	<i>Gryllotalpa</i> sp.	x	GAWZ00000000		-
Mantophasmatodea	<i>Ceuthophilus</i> sp.	x	GAUX00000000		-
	<i>Tetrix subulata</i>	x	GASQ00000000		+
	<i>Prosarthria teretrirostris</i>	x	GAZT00000000		-
	<i>Teleogryllus commodus</i>	x	GBHB00000000		-
	<i>Gryllus firmus</i>	x	GAIZ00000000		-
	<i>Tanzaniophasma</i> sp.	x	GAXB00000000		+
	<i>Zorotypus caudelli</i>	x	GAYA00000000		+
Zoroptera	<i>Zorotypus gurneyi</i>	x	GABA00000000		-

Fortsetzung I Tabelle A 1

Taxon (Klasse/Ordnung)	untersuchte Art	genomisch	transkriptomisch	gen.	trans.
Blattodea	<i>Blattella germanica</i>	PRJNA203136	GBID00000000	+	-
	<i>Periplaneta americana</i>	x	GAW500000000		+
	<i>Cryptocercus wrighti</i>	x	GAZN00000000		+
	<i>Blaberus atropos</i>	x	GAYD00000000		-
Isoptera	<i>Zootermopsis nevadensis</i>	PRJNA203242	x	-	
	<i>Mastotermes darwiniensis</i>	x	GAZE00000000		-
	<i>Prorhinotermes simplex</i>	x	GASE00000000		-
Mantodea	<i>Metallyticus splendidus</i>	x	GATB00000000		+
	<i>Mantis religiosa</i>	x	GASW00000000		-
	<i>Empusa pennata</i>	x	GAWT00000000		-
Dermaptera	<i>Apachyus chartaceus</i>	x	GAUW00000000		-
	<i>Forficula auricularia</i>	x	GAYQ00000000 GAAX00000000		-
Grylloblattodea	<i>Grylloblatta bifratrilecta</i>	x	GAWP00000000		-
	<i>Galloisiana yuasai</i>	x	GAWN00000000		-
Psocoptera	<i>Ectopsocus briggsi</i>	x	GAPT00000000		-
	<i>Liposcelis bostrychophila</i>	x	GAYV00000000		-
Phthiraptera	<i>Pediculus humanus</i>	PRJNA19807	PRJNA19807	+	+
	<i>Menopon gallinae</i>	x	GAWR00000000		+
Thysanoptera	<i>Frankliniella occidentalis</i>	PRJNA203209	x	-	
	<i>Megalurothrips sjostedti</i>	x	GAJY00000000		-
	<i>Gynaikothrips ficorum</i>	x	GAXG00000000		-
	<i>Thrips palmi</i>	x	GAXC00000000		-
	<i>Frankliniella cephalica</i>	x	GAYE00000000		-
	<i>Frankliniella occidentalis</i>	x	GAXD00000000		-
Hemiptera	<i>Bemisia tabaci</i>	x	GAUC00000000 GARP00000000 GARQ00000000 GAPP00000000 GAPQ00000000		-
	<i>Trialeurodes vaporariorum</i>	x	GAWX00000000		-
	<i>Acyrtosiphon pisum</i>	PRJNA29489	PRJNA29489	-	-
	<i>Dactylopius coccus</i>	PRJNA244295	x	-	
	<i>Kerria lacca</i>	x	GBDY00000000 GBDP00000000 GBDO00000000		-
	<i>Macrosiphum euphorbiae</i>	x	GAOM01000000		-
	<i>Adelges tsugae</i>	x	GBJX00000000		-
	<i>Essigella californica</i>	x	GAZF00000000		-
	<i>Aphis craccivora</i>	x	GAJW00000000		-
	<i>Sitobion avenae</i>	x	GAPL00000000		-
Hemiptera	<i>Planococcus citri</i>	x	GAXF00000000		-
	<i>Pachypsylla venusta</i>	PRJNA167476	GAOP00000000	-	-
	<i>Diaphorina citri</i>	PRJNA251515	GACJ00000000	-	-
	<i>Acanthosuarina muellerianae</i>	x	GAYY00000000		-
	<i>Homalodisca vitripennis</i>	PRJNA168119	x	-	
	<i>Okanagana villosa</i>	x	GAWQ00000000		-
	<i>Cercopis vulnerata</i>	x	GAUN00000000		-
	<i>Graminella nigrifrons</i>	x	GAQX00000000		-
	<i>Piezodorus guildinii</i>	PRJNA263369	x	-	
	<i>Gerris buenoi</i>	PRJNA203045	x	-	
	<i>Halyomorpha halys</i>	PRJNA168118	GDCC00000000 GBHT00000000		-
	<i>Oncopeltus fasciatus</i>	PRJNA229125	x	-	
	<i>Cimex lectularius</i>	PRJNA167477	x	-	
	<i>Nilaparvata lugens</i>	PRJNA177647	GAYF00000000 GANM00000000		-
	<i>Rhodnius prolixus</i>	PRJNA13648	GAHY00000000		-

I  
n  
s  
e  
c  
t  
a

Fortsetzung II Tabelle A 1

Taxon (Klasse/Ordnung)	untersuchte Art	genomisch	transkriptomisch	gen.	trans.
Hemiptera	Lygus hesperus	x	GBRD00000000 GBHO00000000	-	-
	Notostira elongata	x	GASV00000000	-	-
	Lygus lineolaris	x	GDAW00000000	-	-
	Rhodnius neglectus	x	GDKW00000000	-	-
	Panstrongylus megistus	x	GBGD00000000	-	-
	Triatoma infestans	x	GBBI00000000	-	-
	Euschistus heros	x	GBER00000000	-	-
	Chinavia ubica	x	GBFA00000000	-	-
	Dichelops melacanthus	x	GBES00000000	-	-
	Acanthosoma haemorrhoidale	x	GAUV00000000	-	-
	Anoplocnemis curvipes	x	GAJV00000000	-	-
	Clavigralla tomentosicollis	x	GAJX00000000	-	-
	Ranatra linearis	x	GAYZ00000000	-	-
	Velia caprai	x	GAUO00000000	-	-
	Xenophysella greensladeae	x	GAYI00000000	-	-
Megaloptera	Corydalinae sp.	x	GADH01000000	-	-
	Sialis lutaria	x	GABK00000000	-	-
	Corydalus cornutus	x	GATG00000000	-	-
Raphidioptera	Raphidia ariadne	x	GACX00000000	-	-
	Xanthostigma xanthostigma	x	GAUI00000000	-	-
Neuroptera	Inocellia crassicornis	x	GAZH00000000	+	-
	Chrysopa pallens	x	GAGF00000000	-	-
	Nevrorthus apatelios	x	GACU00000000	-	-
	Conwentzia psociformis	x	GAYH00000000	-	-
	Osmylus fulvicephalus	x	GAYC00000000	+	-
	Euroleon nostras	x	GAXW00000000	-	-
	Pseudomallada prasinus	x	GAVV00000000	-	-
	Hypothenemus hampei	PRJNA279497	x	+	-
	Anoplophora glabripennis	PRJNA167479	SRR1799852	+	+
	Agrilus planipennis	PRJNA230921	SRR1791269	+	+
Coleoptera	Leptinotarsa decemlineata	PRJNA171749	SRR1827566	+	+
	Onthophagus taurus	PRJNA167478	SRR1864900	+	+
	Triboleum castaneum	PRJNA15718	SRR1647926/36	+	+
	Dendroctonus ponderosae	PRJNA179493	SRR867432-41	+	+
	Dastarcus helophoroides	x	GBCX00000000	+	-
	Ips typographus	x	GACR01000000	-	-
	Nicrophorus vespilloides	x	GDKQ00000000	-	-
	Meligethes aeneus	x	GAPE00000000	-	-
	Priacma serrata	x	GACO01000000	-	-
	Carabus granulatus	x	GACW00000000	-	-
	Aethina tumida	x	GCKB01000000	+	-
	Popillia japonica	x	GARK00000000	-	-
	Colaphellus bowringi	x	GBHN00000000	+	-
	Lepicerus sp.	x	GAZB00000000	+	-
	Meloe violaceus	x	GATA00000000	-	-
	Gyrinus marinus	x	GAUY00000000	-	-
	Aleochara curtula	x	GATW00000000	+	-
	Onthophagus nigriventris	x	GAQW00000000	-	-
	Trypoxylus dichotomus	x	GAQV00000000	-	-
	Anthonomus grandis	x	GABY00000000	+	-
Dendroctonus frontalis	x	GAFI01000000	-	-	
Pissodes strobi	x	GAEO00000000	-	-	
Strepsiptera	Mengenilla moldrzyki	PRJNA72521	GACY00000000	-	-
	Stylops melittae	x	GAZM00000000	-	-

Fortsetzung III Tabelle A 1

Taxon (Klasse/Ordnung)	untersuchte Art	genomisch	transkriptomisch	gen.	trans.
Mecoptera	Nannochorista philpotti	x	GADB00000000	-	-
	Panorpa vulgaris	x	GAUH00000000	-	-
	Boreus hyemalis	x	GAYK00000000	-	-
	Bittacus pilicornis	x	GATH00000000	-	-
Siphonaptera	Ctenocephalides felis	x	GAYP00000000	-	-
	Ceratophyllus gallinae	x	GAWK00000000	-	-
	Archaeopsylla erinacei	x	GABB00000000	-	-
	Oropsylla silantiewi	x	GAWY00000000	-	-
Trichoptera	Limnephilus lunatus	PRJNA203303	x	-	-
	Philopotamus ludificatus	x	GACV00000000	-	-
	Platycentropus radiatus	x	GASS00000000	-	-
	Annulipalpia sp.	x	GATX00000000	-	-
	Rhyacophila fasciata	x	GAXX00000000	-	-
	Hydroptila sp.	x	GAVM00000000	-	-
Lepidoptera	Heliconius melpomene	PRJEA71053	x	-	-
	Bombyx mori	PRJNA205630	GBJR01000000	-	-
	Papilio glaucus	PRJNA270125	x	-	-
	Melitaea cinxia	PRJNA191594	x	-	-
	Papilio polytes	PRJNA291535	PRJNA291535	-	-
	Amyelois transitella	PRJNA292025	PRJNA292025	-	-
	Papilio xuthus	PRJNA291600	PRJNA291600	-	-
	Plutella xylostella	PRJNA277936	PRJNA277936	-	-
	Athetis lepigone	x	GARD00000000 GARG00000000 GARC00000000 GARB00000000	-	-
	Spodoptera exigua	x	GARL00000000 GAOR00000000 GAOQ00000000 GAFU00000000	-	-
	Thaumetopoea pityocampa	x	GBY00000000 GBYZ00000000 GBZB00000000	-	-
	Lymantria dispar	x	GDCN00000000 GDDM00000000	-	-
	Ostrinia nubilalis	x	GAVD00000000 GAHP00000000	-	-
	Helicoverpa assulta	x	GBTA00000000	-	-
Helicoverpa armigera	x	GBDM00000000	-	-	
Agrotis segetum	x	GBCW00000000	-	-	
Spodoptera litura	x	GBBY00000000	-	-	
Spodoptera frugiperda	x	GCTM00000000	-	-	
Trichoplusia ni	x	GBKU00000000	-	-	
Ostrinia scapulalis	x	GAHQ00000000	-	-	
Ostrinia furnacalis	x	GAQJ00000000	-	-	
Chilo suppressalis	x	GAJS00000000	-	-	
Amyelois transitella	x	GDGN00000000	-	-	
Cadra cautella	x	GBXH00000000	-	-	
Antheraea pernyi	x	GBZF00000000	-	-	
Antheraea yamamai	x	GBZJ00000000	-	-	
Antheraea assama	x	GBZC00000000	-	-	
Actias selene	x	GBZL00000000	-	-	
Plutella xylostella	x	GBAC00000000	-	-	
Yponomeuta evonymellus	x	GASG00000000	-	-	
Grapholita dimorpha	x	GADJ00000000	-	-	
Grapholita molesta	x	GADK00000000	-	-	
Telchin licus	x	GBAS00000000	-	-	
Zygaena fausta	x	GAYB00000000	-	-	
Carposina sasakii	x	GADL00000000	-	-	

I  
n  
s  
e  
c  
t  
a

Fortsetzung IV Tabelle A 1

Taxon (Klasse/Ordnung)	untersuchte Art	genomisch	transkriptomisch	gen.	trans.
Lepidoptera	<i>Samia ricini</i>	x	GBZD00000000	-	-
	<i>Rhodinia newara</i>	x	GBZE00000000	-	-
	<i>Polyommatus icarus</i>	x	GAST00000000	-	-
	<i>Parides eurimedes</i>	x	GAXH00000000	-	-
	<i>Pararge aegeria</i>	x	GAXI00000000	-	-
	<i>Nemophora degeerella</i>	x	GATC00000000	-	-
	<i>Triodia sylvina</i>	x	GAVB00000000	-	-
	<i>Dyseriocrania subpurpurella</i>	x	GASY00000000	-	-
	<i>Micropterix calthella</i>	x	GACP00000000	-	-
Hymenoptera	<i>Apis mellifera</i>	PRJNA13343	GALO00000000 GAZV00000000	-	-
	<i>Apis florea</i>	PRJNA86991	PRJNA86991	-	-
	<i>Apis dorsata</i>	PRJNA232132	PRJNA232132	-	-
	<i>Nasonia vitripennis</i>	PRJNA20073	GBEB00000000	-	-
	<i>Nasonia giraulti</i>	PRJNA20223	GBEC00000000	-	-
	<i>Nasonia longicornis</i>	PRJNA20225	x	-	-
	<i>Pogonomyrmex barbatus</i>	PRJNA276107	PRJNA276107	-	-
	<i>Linepithema humile</i>	PRJNA281950	PRJNA281950	-	-
	<i>Atta cephalotes</i>	PRJNA279976	PRJNA279976	-	-
	<i>Camponotus floridanus</i>	PRJNA274144	PRJNA274144	-	-
	<i>Harpegnathos saltator</i>	PRJNA273397	PRJNA273397	-	-
	<i>Solenopsis invicta</i>	PRJNA268798	PRJNA268798	-	-
	<i>Bombus terrestris</i>	PRJNA68545	PRJNA68545	-	-
	<i>Bombus impatiens</i>	PRJNA70395	PRJNA70395	-	-
	<i>Acromyrmex echinator</i>	PRJNA271903	PRJNA271903	-	-
	<i>Megachile rotundata</i>	PRJNA87021	PRJNA87021	-	-
	<i>Cephus cinctus</i>	PRJNA168335	x	-	-
	<i>Athalia rosae</i>	PRJNA282653	PRJNA282653	+	+
	<i>Ceratosolen solmsi</i>	PRJNA277475	PRJNA277475	-	-
	<i>Orussus abietinus</i>	PRJNA282746	GAUJ00000000	-	-
	<i>Microplitis demolitor</i>	PRJNA251518	GANH00000000	-	-
	<i>Wasmannia auropunctata</i>	PRJNA279179	PRJNA279179	-	-
	<i>Monomorium pharaonis</i>	PRJNA285339	PRJNA285339	-	-
	<i>Vollenhovia emeryi</i>	PRJNA278668	PRJNA278668	-	-
	<i>Trichogramma pretiosum</i>	PRJNA168121	x	-	-
	<i>Cerapachys biroi</i>	PRJNA275884	x	-	-
	<i>Copidosoma floridanum</i>	PRJNA171748	x	-	-
	<i>Fopius arisanus</i>	PRJNA274979	GBYB00000000 SRR1560649/53	+	+
	<i>Cotesia vestalis</i>	PRJNA271135	GAKG00000000 GAUP00000000	-	-
	<i>Lasius niger</i>	PRJNA269328	x	-	-
	<i>Polistes metricus</i>	x	GDHQ00000000 GBGV00000000	-	-
	<i>Polistes fuscatus</i>	x	GDFS00000000	-	-
	<i>Polistes canadensis</i>	x	GAFR00000000	-	-
<i>Mischocyttarus flavitarsis</i>	x	GAXM00000000	-	-	
<i>Pseudomasaris vespoides</i>	x	GAXQ00000000	-	-	
<i>Chyphotes mellipes</i>	x	GAXL00000000	-	-	
<i>Sphaerophthalma orestes</i>	x	GAXP00000000	-	-	
<i>Tetramorium bicarinatum</i>	x	GASM00000000	-	-	
<i>Dinoponera quadriceps</i>	x	GANS00000000	-	-	
<i>Stigmatomma oregonense</i>	x	GAXR00000000	-	-	
<i>Brachycistis timberlakei</i>	x	GAZU00000000	-	-	
<i>Pepsis grossa</i>	x	GAXS00000000	-	-	
<i>Crioscolia alcione</i>	x	GAXT00000000	-	-	
<i>Sceliphron caementarium</i>	x	GAXN00000000	-	-	
<i>Argochrysis armilla</i>	x	GAXO00000000	-	-	

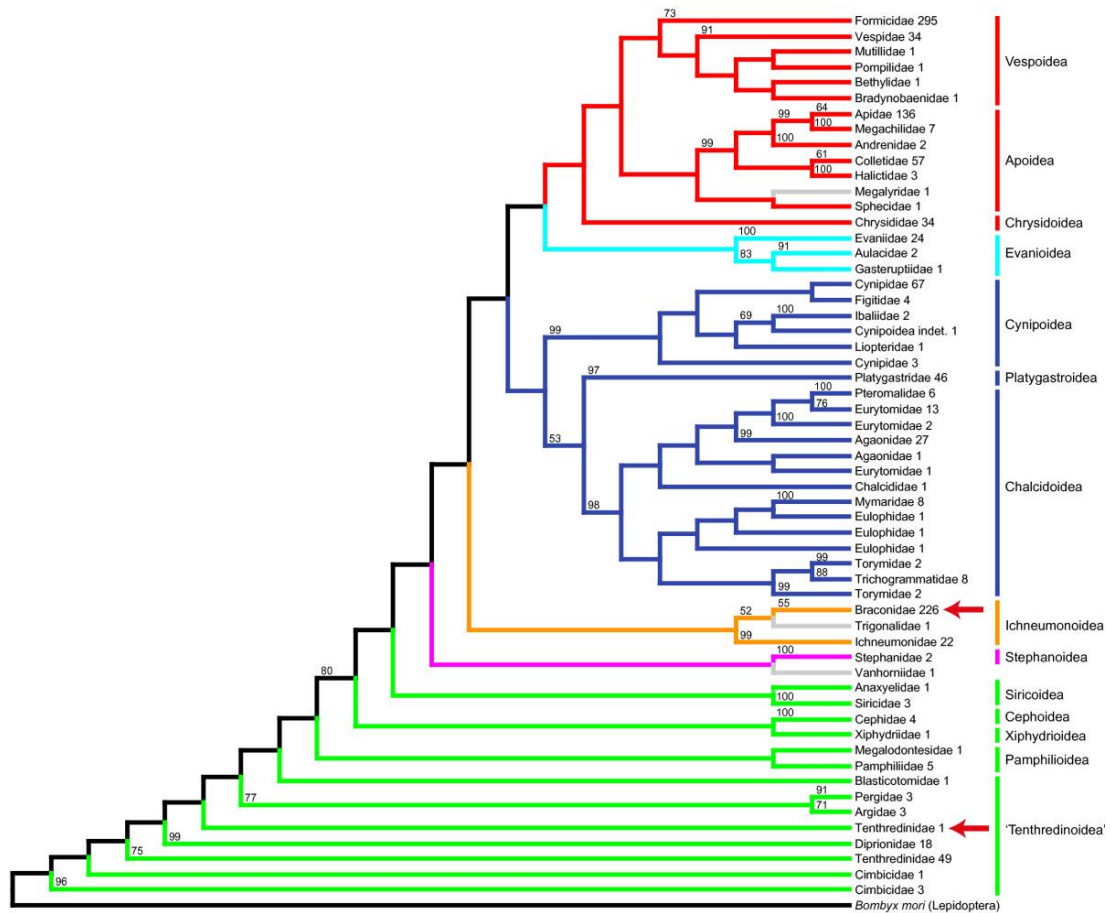
Fortsetzung V Tabelle A 1

Taxon (Klasse/Ordnung)	untersuchte Art	genomisch	transkriptomisch	gen.	trans.
Hymenoptera	<i>Chrysis viridula</i>	x	GATY00000000	-	-
	<i>Telenomus podisi</i>	x	GBEU00000000	-	-
	<i>Leptopilina bouardi</i>	x	GAJA00000000	-	-
	<i>Leptopilina heterotoma</i>	x	GAJC00000000	-	-
	<i>Leptopilina clavipes</i>	x	GAXY00000000	-	-
	<i>Ganaspis</i> sp.	x	GAIW00000000	-	-
	<i>Osmia cornuta</i>	x	GAGH00000000	-	-
	<i>Xyela alpigena</i>	x	GADA00000000	-	-
	<i>Tenthredo koehleri</i>	x	GAWW00000000	-	-
Diptera	<i>Aedes aegypti</i>	PRJNA19731	PRJNA19731	-	-
	<i>Anopheles gambiae</i>	PRJNA163	PRJNA163	-	-
	<i>Culex quinquefasciatus</i>	PRJNA29017	PRJNA29017	-	-
	<i>Chironomus tentans</i>	PRJEB1888	x	-	-
	<i>Chironomus riparius</i>	PRJNA268391	x	-	-
	<i>Belgica antarctica</i>	PRJNA172148	GAAK00000000	-	-
	<i>Chaoborus trivitattus</i>	PRJNA268391	x	-	-
	<i>Mochlonyx cincipes</i>	PRJNA268391	x	-	-
	<i>Anopheles sinensis</i>	x	GAFE00000000 GBEO00000000	-	-
	<i>Psorophora albipes</i>	x	GALA00000000	-	-
	<i>Anopheles aquasalis</i>	x	GAMD00000000	-	-
	<i>Corethrella appendiculata</i>	x	GANO00000000	-	-
	<i>Aedes albopictus</i>	x	GAPW00000000	-	-
	<i>Culicoides sonorensis</i>	x	GAWM00000000	-	-
	<i>Anopheles quadrimaculatus</i>	x	GBTE00000000	-	-
	<i>Anopheles stephensi</i>	x	GBVY00000000	-	-
	<i>Mayetiola destructor</i>	PRJNA45867	x	-	-
	<i>Sitodiplosis mosellana</i>	x	GAKJ00000000 GBRL00000000	-	-
	<i>Bibio marci</i>	x	GATJ00000000	-	-
	<i>Trichoceridae</i> sp.	PRJNA268391	x	-	-
	<i>Clogmia albipunctata</i>	PRJNA268391	x	-	-
	<i>Coboldia fuscipes</i>	PRJNA268391	x	-	-
	<i>Phlebotomus papatasi</i>	PRJNA20293	x	-	-
	<i>Lutzomyia longipalpis</i>	PRJNA20279	x	-	-
	<i>Trichocera saltator</i>	x	GAXZ00000000	-	-
	<i>Tipula oleracea</i>	PRJNA268391	x	-	-
	<i>Tipula maxima</i>	x	GACZ00000000	-	-
	<i>Drosophila melanogaster</i>	PRJNA164	PRJNA164	-	-
	<i>Musca domestica</i>	PRJNA210139	PRJNA210139 GDAV00000000 GARN00000000	-	-
	<i>Megaselia scalaris</i>	PRJEB1273	x	-	-
	<i>Glossina fuscipes</i>	PRJNA172853	x	-	-
	<i>Bactrocera tryoni</i>	PRJNA241080	x	-	-
	<i>Phortica variegata</i>	PRJNA268391	x	-	-
<i>Scaptodrosophila lebanonensis</i>	PRJNA268391	x	-	-	
<i>Themira minor</i>	PRJNA268391	x	-	-	
<i>Trupanea jonesi</i>	PRJNA268391	x	-	-	
<i>Condylostylus patibulatus</i>	PRJNA268391	x	-	-	
<i>Hermetia illucens</i>	PRJNA268391	x	-	-	
<i>Liriomyza trifolii</i>	PRJNA268391	x	-	-	
<i>Ephydra hians</i>	PRJNA268391	x	-	-	
<i>Eutreta diana</i>	PRJNA268391	x	-	-	
<i>Eristalis dimidiata</i>	PRJNA268391	x	-	-	
<i>Megaselia abdita</i>	PRJNA268391	x	-	-	
<i>Holcocephala fusca</i>	PRJNA268391	x	-	-	

Fortsetzung VI Tabelle A 1

Taxon (Klasse/Ordnung)	untersuchte Art	genomisch	transkriptomisch	gen.	trans.
S f u e s n i Diptera	<i>Sphyracephala brevicornis</i>	PRJNA268391	x	-	
	<i>Calliphora vicina</i>	PRJNA268391	x	-	
	<i>Neobellieria bullata</i>	PRJNA268391	x	-	
	<i>Tephritis californica</i>	PRJNA268391	x	-	
	<i>Teleopsis dalmanni</i>	PRJNA268391	GBBP00000000		-
	<i>Lucilia cuprina</i>	PRJNA203545	x	-	
	<i>Stomoxys calcitrans</i>	PRJNA288986	GDIM00000000		-
	<i>Bactrocera oleae</i>	PRJNA293367	GAKB00000000		-
	<i>Bactrocera minax</i>	x	GBEY00000000		-
	<i>Bactrocera dorsalis</i>	x	GAKP00000000		-
	<i>Bactrocera latifrons</i>	x	GDHF00000000		-
	<i>Bactrocera cucurbitae</i>	x	GBXI00000000		-
	<i>Ceratitis capitata</i>	x	GAMC00000000		-
	<i>Teleopsis whitei</i>	x	GBBQ00000000		-
	<i>Lipara lucens</i>	x	GAZD00000000		-
	<i>Drosophila serrata</i>	x	GAHN00000000		-
	<i>Drosophila malerkotliana malerkotliana</i>	x	GADM00000000		-
	<i>Drosophila malerkotliana pallens</i>	x	GADQ00000000		-
	<i>Drosophila pseudoananassae pseudoananassae</i>	x	GAEJ00000000		-
	<i>Drosophila pseudoananassae nigrens</i>	x	GADR00000000		-
	<i>Drosophila ercepeae</i>	x	GADP00000000		-
	<i>Drosophila merina</i>	x	GADS00000000		-
	<i>Drosophila miranda</i>	x	GALP00000000		-
	<i>Themira biloba</i>	x	GBGG00000000		-
	<i>Triarthria setipennis</i>	x	GAVA00000000		-
	<i>Tabanus bromius</i>	x	GDAI00000000		-
	<i>Bombylius major</i>	x	GATI00000000		-
	<i>Delia antiqua</i>	x	GAWI00000000		-





**Abbildung A 11 | Stammbaum der Ordnung der Hautflügler (*Hymenoptera*).** Dargestellt sind die einzelnen Familien innerhalb der *Hymenoptera* und der Anzahl der je Familie untersuchten Arten. Die Bootstrap-Werte sind neben den jeweiligen Knoten gezeigt, Werte unter 50 % sind ausgelassen. Die übergeordneten Abstammungslinien sind farblich kodiert, Familien mit unsicherer Position sind in grau dargestellt. Die Familien von *Fopius arisanus* (*Braconidae*) und *Athalia rosae* (*Tenthredinidae*) sind durch einen roten Pfeil markiert. (übernommen und verändert nach: Peters, RS; Meyer, B; Krogmann, L; Borner, J; Meusemann, K; Schütte, K et al.: *The taming of an impossible child: a standardized all-in approach to the phylogeny of Hymenoptera using public database sequences*. BMC biology 201155. [256])

## A6 Inhaltsverzeichnis des elektronischen Anhangs

Folgende Dateien befinden sich im beigefügten elektronischen Anhang:

- E1 – verwendete Perl-Skripte:
  - fasta\_slice
- E2 – Primerlisten
- E3 – identifizierte *stk33*-Sequenzen der Klasse *Aves* als gbk-Dateien
- E4 – identifizierte *stk33*-Sequenzen innerhalb des Reiches *Animalia* als gbk-Dateien
- E5 – vorliegende Doktorarbeit im PDF-Format
- E6 – Publikation im PDF-Format

## Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation selbstständig und ohne fremde Hilfe angefertigt zu haben. Weiterhin versichere ich, örtlich übernommene Ausführungen anderer Autoren und an Gedankengänge Anderer anlehrende eigene Formulierungen entsprechend gekennzeichnet und die Quellen zitiert zu haben.

-----  
(Ort, Datum)

-----  
(Tobias Lautwein)