Developmental genetics of *Xenopus*

By

Jessica Burnham Lyons

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Richard Harland, chair
Professor Sharon Amacher
Professor John Gerhart
Professor Gregory Aponte

Spring 2012

Developmental genetics of *Xenopus*

By Jessica Burnham Lyons

Abstract

Developmental genetics of *Xenopus*

by

Jessica Burnham Lyons

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Richard Harland, Chair

The frog *Xenopus laevis* has been studied for over 200 years. Its advantages as a robust and easily manipulable model organism have been complemented more recently by genetic and genomic studies in *Xenopus tropicalis*. My thesis work utilized established and cutting-edge techniques to advance our knowledge of developmental genetics in both *Xenopus* systems.

I used an embryological approach to investigate the roles of the Fgf receptors (Fgfrs) during development. Knockdown and overexpression studies suggested that each Fgfr plays a different role in the specification of mesoderm, and my results are consistent with Fgfr4 playing a role in dampening the Fgf signal. I also showed that *X. tropicalis* Fgfrs 1–3 are alternatively spliced in D3, the extracellular immunoglobulin domain important for ligand specificity. These isoforms exhibit different temporal and spatial expression patterns, suggesting that control of this alternative splicing plays a role in regulating development.

My thesis work has also harnessed the power of the *X. tropicalis* system to understand development using a forward genetic approach. Tadpoles homozygous for the recessive lethal mutation *curly* exhibit ventral edema and curled tails. I used classical genetics to map *curly* to a 1.9 Mb window on *X. tropicalis* chromosome 4. The *pteg* gene, which lies in this region, is misspliced in *curly* embryos. Isabelle Philipp and I used next-generation (nextgen) sequencing technology to identify differences between the *curly* mutant DNA and that of the reference genome. Focusing on the region around the *pteg* gene, we are currently evaluating these differences to find the lesion that causes the *curly* phenotype.

The efficient application of genetic techniques requires a high-quality genetic map and reference genome, and those available for *X. tropicalis* were flawed (Wells et al., 2011) (Hellsten et al., 2010). Thus, I developed a genotyping by sequencing (GBS) technique using reduced-representation multiplex nextgen sequencing, and my collaborators in the Rokhsar lab have generated a high-quality SNP map for *X. tropicalis* based on my data from 192 F2 individuals. This map is being used to construct an improved *X. tropicalis* genome assembly. By finding regions of correlation for the pigmentation mutation *gray*, as well as the *X. tropicalis* sex locus, we demonstrated that this technique can be used for genetic mapping. I

applied my GBS method to *X. laevis* as well, and the SNP map thus generated will facilitate the assembly of a reference genome for this allotetraploid species.

By combining the classical advantages of the *Xenopus* system with modern techniques, my thesis work has contributed to our understanding of the development, genetics, and genomics of vertebrate biology.

Table of Contents

Introduction

The South African clawed frog *Xenopus laevis* has been studied since the early nineteenth century, and became an important organism in the study of endocrinology in the 1930's (Gurdon and Hopwood, 2000). Hogben and others showed that injection of the urine of pregnant women into *X. laevis* females was a robust assay for human pregnancy (Crew, 1939), and these animals entered laboratories around the world. Due to its general hardiness and its ability to respond to human chorionic gonadotropin, *X. laevis* became a favorite choice of biochemical embryologists, who needed large numbers of synchronized embryos for their experiments. In the 1950's, Pieter Nieuwkoop played a pivotal role in establishing *X. laevis* as a model for embryological study (Gurdon and Hopwood, 2000). He produced the Normal Table describing its developmental stages, still in use today (Nieuwkoop, 1994).

X. laevis has many advantages as a model organism: it undergoes external fertilization, and hundreds of large (~1.3 mm) eggs are easily obtained at any time of the year via hormone injection (Sive, 2000). The embryos can be cultured in a simple saline solution and are easily manipulated; the distribution of yolk into all embryonic cells facilitates explant and "cut and paste" studies (Sive, 2000). Work in *Xenopus* continues to make important contributions to the fields of biochemistry, neurobiology, cell biology, and developmental biology (Khokha, 2012), and *Xenopus* structures including the tail and the limb have become models in the burgeoning field of regenerative research (Slack et al., 2008). As an amphibian, *Xenopus* is phylogenetically placed between the amniotes and teleost fish, and recent work has shown that large genomic regions are syntenic between frog, chicken, and human (Hellsten et al., 2010). My thesis research on *Xenopus* embodies and has advanced the possibilities for developmental, genetic, and genomic research using this system.

Research on *Xenopus* has entered the genomic era with the development of the closely related *X. tropicalis* as a genetic model organism (Amaya et al., 1998) (Abu-Daya et al., 2012). The lineages leading to *X. laevis* and *X. tropicalis* diverged approximately 50–65 million years ago (Evans et al., 2004), and subsequently the lineage leading to *X. laevis* underwent allotetraploidization (Hellsten et al., 2007). Both copies of an estimated 25 to 50% of genes have been retained in *X. laevis* (Hellsten et al., 2007), and the presence of these extra gene copies, called alloalleles or homeologues, can impede genetic studies. *X. tropicalis*, on the other hand, is diploid but retains almost all of the advantages of the *X. laevis* system, with smaller embryos (~0.7 mm egg size) but a faster generation time (approximately six months to sexual maturity, versus about a year for *X. laevis*). The *X. laevis* genome is 3.1 gigabases (Gb) in size on 18 chromosomes, whereas the *X. tropicalis* genome is only approximately 1.7 Gb on 10 chromosomes (Tymowska and Fischberg, 1973) (Hellsten et al., 2010). Thus, *X. tropicalis* was chosen as the first amphibian genome to be sequenced. Paired plasmid, fosmid, and bacterial artificial chromosome (BAC) ends were Sanger sequenced, and an average coverage of 7.68X was achieved; 97.6% of known genes were present in *X. tropicalis* genome assembly version 4 (Hellsten et al., 2010).

As part of the development of *X. tropicalis* as a genetic model organism and to complement the genome project, *Xenopus* researchers have made a concerted effort to sequence a large number of expressed sequence tags (ESTs) from various developmental stages of *X.*

tropicalis. In particular, Michael Gilchrist has led efforts to combine multiple EST sequences into so-called clusters in order to define the full-length transcript of a gene (Gilchrist et al., 2004). Over 1.2 million *X. tropicalis* ESTs were utilized for the annotation of gene models in the genome assembly (Hellsten et al., 2010).

Injection of mRNAs into *Xenopus* embryos has greatly advanced our understanding of the signaling and patterning that occurs during development (Khokha, 2012). Injection of antisense oligo deoxynucleotides into oocytes (Torpey et al., 1992), or more recently, antisense morpholino oligonucleotides (MOs) (Heasman, 2002), into embryos, has facilitated knockdown studies in *Xenopus*. As part of my thesis work, I used MOs as well as mRNA overexpression to address the roles of the Fgf receptors in mesoderm specification. I took advantage of the *X. tropicalis* genome and associated EST sequences to identify previously-uncharacterized alternatively spliced isoforms of the Fgf receptors. Since then, a method for gene knockdown in *Xenopus* embryos using RNAi has been reported (Lund et al., 2011), as has the generation of targeted mutations in the *X. tropicalis* genome using zinc finger nucleases (Young et al., 2011). More efficient methods for transgenesis continue to be elucidated, including those that take advantage of the Gal4-UAS and Cre-Lox systems (Love et al., 2011), and the recapitulation of endogenous gene expression from a bacterial artificial chromosome (BAC) (Fish et al., 2011).

Taking advantage of the *X. tropicalis* genome project, a set of over 2800 microsatellite markers was generated bioinformatically, and these markers have facilitated genetic mapping in *X. tropicalis* (Abu-Daya et al., 2009; Khokha et al., 2009; Geach and Zimmerman, 2010; Wells et al., 2011). When I began mapping the *curly* gene, a recessive mutation that exhibits Mendelian inheritance, efforts to map it had lasted over four years (Khokha et al., 2009). By using version 7 of the *X. tropicalis* genome assembly, with its chromosome-scale scaffolds, to order the markers, I was able to map the *curly* mutation from a chromosome arm down to a 1.9 megabase (Mb) window in less than a year. By then, the length and number of reads that could be obtained via next-generation (nextgen) sequencing had so advanced that Isabelle Philipp and I were able to obtain and analyze approximately 20X coverage of the genome from mutant embryos from one lane of sequencing. Making the sequencing library, sequencing it, and analyzing the data took approximately four months, and Isabelle and I are currently using these data to identify the mutation. It is reasonable to think that in the not-so-distant future, especially if isogenetic frogs are mutagenized, mutations can rapidly be identified in this manner without the need for laborious and time-consuming genetic mapping. Thus, the *curly* project demonstrates the pace at which new technologies are transforming our approaches to *Xenopus* genetics.

Nonetheless, the incomplete nature of the *X. tropicalis* genome assembly and the demonstrated lack of resolution in the existing microsatellite-based genetic map presented obstacles to efficient genetic mapping in this species (Abu-Daya et al., 2009; Hellsten et al., 2010; Wells et al., 2011). For this reason, I set out to construct a high-density genetic map based on clearly distinguishable single nucleotide polymorphisms (SNPs), with the goal of mitigating these issues. In collaboration with Therese Mitros, a member of the Rokhsar lab at UC Berkeley, I used next-generation sequencing to generate an SNP-based genetic map of the *X. tropicalis* genome. This map is being used to prepare an authoritative *X. tropicalis*

genome assembly, and Therese and I have also shown that the reduced-representation multiplex approach I developed can be used to identify regions of correlation for mutations or traits.

In the years since the *X. tropicalis* genome project was begun, Sanger sequencing has become dispensable for genome analysis, in large part due to the availability of next-generation (nextgen) sequencing technologies that can generate large amounts of data at minmal expense (Metzker, 2010). Indeed, the Rokhsar lab has developed a technique for genome assembly *de novo* from nextgen sequencing data (Chapman et al., 2011), and they are applying this technique to the construction of an *X. laevis* genome assembly (D. Rokhsar, personal communication). To complement their efforts, my collaborator and I are generating a genetic map for *X. laevis* using the genotyping by sequencing approach I developed for *X. tropicalis*. This map will help to overcome the hurdle of differentiating chromosomes containing homeologous genes from one another, once thought to be a technical challenge to *X. laevis* genome assembly. Furthermore, the strain-specific genome sequences I have generated for *X. tropicalis* and *X. laevis* represent the future of genomics, in which there will no longer be one reference genome sequence for each species. Indeed, the genomes of 17 mouse strains have recently been sequenced via nextgen sequencing (Keane et al., 2011), and over 700,000 structural variations (such as retrotransposons and various inserted or deleted repeats) were found between them, which in rare cases had a large effect on gene function (Yalcin et al., 2011).

My thesis research has exploited and developed tools that demonstrate the utility of *Xenopus* as a model organism. It also exemplifies the fast pace at which molecular biology research advances, particularly in the fertile collaborative environment of UC Berkeley.

Acknowledgements

I would like to thank Richard and the members of my thesis committee for support, guidance, and feedback.

It has been a pleasure working with my collaborators Therese Mitros, Dan Rokhsar, and Isa Philipp.

Members of the Harland lab, past and present, I could not have made it through grad school without you and I am honored to have been in your company.

To my family and friends, thank you so much for your support, patience, and understanding over the past six years.

Chapter One: The roles of the Fgf receptors in *Xenopus* mesoderm specification

I. Introduction

Fibroblast growth factors (Fgfs) are an animal-specific family of signaling molecules first identified by their ability to stimulate the growth of fibroblasts in culture (Gospodarowicz, 1974). Fgf signaling is involved in diverse developmental processes in vertebrates (Dorey and Amaya, 2010), including induction of mesoderm (Kimelman and Kirschner, 1987; Slack et al., 1987; Kimelman et al., 1988; Amaya et al., 1991; Amaya et al., 1993), neural patterning (Lamb and Harland, 1995; Monsoro-Burq, 2003), convergent extension (Nutt et al., 2001; Frazzetto et al., 2002; Chung et al., 2004), and limb and skeletal development (Ornitz and Marie, 2002).

The first studies on the role of Fgfs in *Xenopus laevis* showed that they can induce ectodermal explants (animal caps) to adopt a mesodermal fate (Kimelman and Kirschner, 1987; Slack et al., 1987; Kimelman et al., 1988). The injection of RNA encoding XFD, a dominant-negative form of Fgfr1, results in embryos that are severely lacking in trunk and tail mesoderm and are posteriorly truncated (Amaya et al., 1991; Amaya et al., 1993). Fgf signaling is required for the proper expression of the early mesodermal marker *t* (formerly known as *XBra* in *Xenopus*) (Amaya et al., 1993; Isaacs et al., 1994; Fletcher and Harland, 2008) as well as muscle markers such as *myoD* and *myf5* (Standley et al., 2001; Fisher et al., 2002).

The *X. tropicalis* Fgf family comprises 20 ligands numbered 1-14, 16, 19, 20, 22, 23.1, and 23.2, based on similarity at the DNA sequence level, as well as synteny, to the human Fgfs (xenbase.org). A number of Fgf ligands have been implicated in *Xenopus* mesoderm specification, notably Fgf3, 4, 8b, and 9, which are all expressed in the presumptive mesodermal domain around the blastopore during gastrula stages, and can induce mesoderm in animal cap explants (Isaacs et al., 1992; Song and Slack, 1996; Lombardo et al., 1998; Fletcher et al., 2006).

Fgf ligands bind to their cognate receptors in a 2:2:2 complex between ligands, heparan sulfate proteoglycans (HSPGs), and dimerized receptors (Figure 1.1) (Schlessinger et al., 2000). Fgf receptors (Fgfrs) are receptor tyrosine kinases that undergo transphosphorylation upon ligand binding (Böttcher and Niehrs, 2005). Their domain structure consists of three extracellular immunoglobulin-like (Ig) domains (also referred to as D1-D3), a transmembrane region, and an intracellular tyrosine kinase domain (Figure 1.1) (Givol and Yayon, 1992). D2 and D3 participate in ligand binding (Chellaiah et al., 1999), and ligand specificity is determined by D3 (Yayon et al., 1992; Plotnikov et al., 2000). Mammalian Fgfrs 1–3 are known to undergo alternative splicing of mutually exclusive exons encoding the C-terminal half of D3 (Figure 1.10A) (Eswarakumar et al., 2005). The resulting spliceforms, referred to as IIIb and IIIc, differ in ligand binding specificity *in vitro* and are sometimes expressed in different tissue types (Miki et al., 1992; Avivi et al., 1993; Orr-Urtreger et al., 1993; Yan et al., 1993; Chellaiah et al., 1994; Plotnikov et al., 2000). Intracellular transduction of the Fgf signal can occur three ways: through the Ras/MAPK, PI3K/Akt, or PLCγ pathways (Eswarakumar et al., 2005). Both the MAPK and PI3K

pathways are involved in Fgf signaling-mediated mesoderm induction (MacNicol et al., 1993; Carballada et al., 2001). The MAPK pathway is also required for Fgf-mediated neural posteriorization (Ribisi et al., 2000), and the Fgf signal is transduced into morphogenetic movements via the PLCγ pathway (Nutt et al., 2001).

Like other vertebrates, *Xenopus* has four Fgf receptors (Givol and Yayon, 1992). Previous studies on the role of Fgf signaling during *Xenopus* development focused on specific ligands or used general inhibitors of Fgf signaling such as the dominant-negative receptor XFD or the small molecule inhibitor SU5402 (Amaya et al., 1991; Amaya et al., 1993; Chung et al., 2004; Delaune et al., 2005; Fletcher and Harland, 2008). As noted above, such studies have shown that Fgf signaling is required for proper mesoderm induction, including the specification of trunk and tail mesoderm, and proper initiation and maintenance of *t* expression. The contribution, however, of each Fgf receptor to this process is largely unknown in any species. I addressed this question using an antisense morpholino oligonucleotide (MO) knockdown approach, and I found that MOs designed against Fgfrs 1, 3, and 4 cause a dose-dependent decrease in *t* expression at gastrula stages, suggesting that these genes are required for proper mesoderm specification. Fgfr overexpression, however, did not robustly increase *t* expression, and in the case of Fgfr4, it caused a decrease in *t* expression in the marginal zone, suggesting that Fgf-mediated mesoderm specification is tightly regulated. Experiments to determine the character of marginal zone cells lacking *t* expression when Fgf signaling is manipulated were inconclusive.

Before I undertook the work described here, it was also not known whether the *Xenopus* Fgf receptors are alternatively spliced in the D3 domain nor what roles the different isoforms play in the regulation of development. By taking advantage of the *X. tropicalis* genome assembly and large collection of ESTs, I identified the D3 alternative spliceforms of Fgfrs 1, 2, and 3. RT–PCR and *in situ* hybridization revealed that these isoforms are differentially expressed.

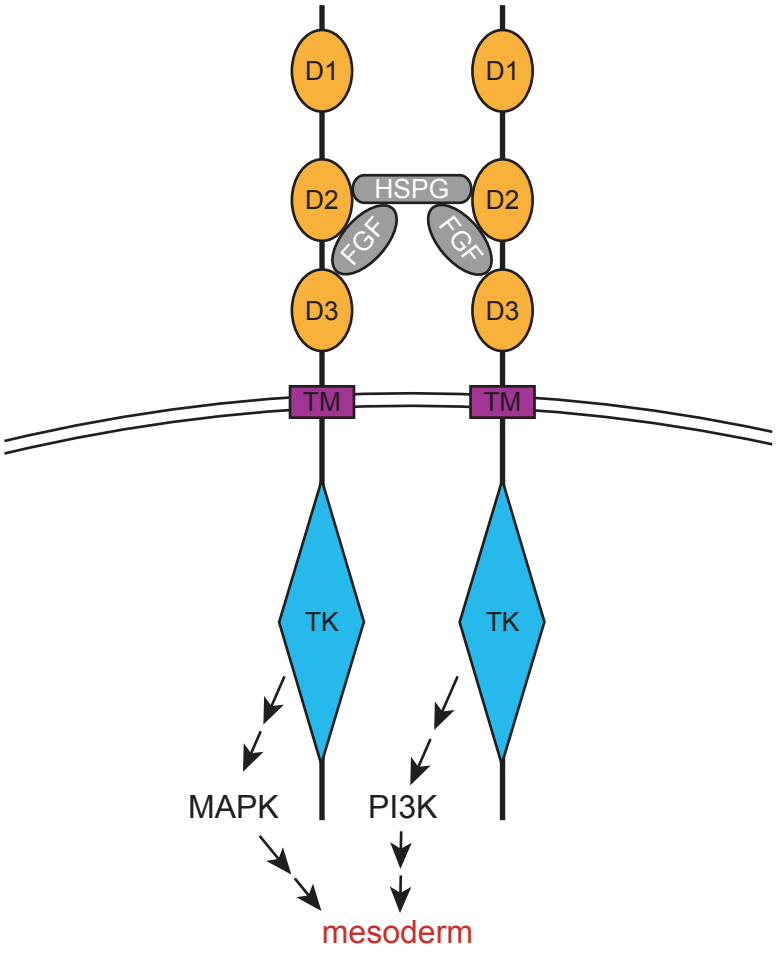


Figure 1.1. Domain structure of Fgf receptors. Fgfrs have three extracellular immunoglobulin domains (referred to as D1–D3), a transmembrane domain, and an intracellular tyrosine kinase domain (Givol and Yayon, 1992). Fgf ligands bind their cognate receptors in a 2:2 complex that requires heparin sulfate proteoglycans (HSPGs) (Schlessinger et al., 2000). Both the MAPK and PI3K pathways are involved in Fgf signaling-mediated mesoderm induction (MacNicol et al., 1993; Carballada et al., 2001).

II. Fgfr knockdown and overexpression studies

IIa. Morpholino knockdown

In order to parse the individual contributions of the Fgfrs to mesoderm specification, I sought to test whether each is required for this process by individually knocking down the expression of each Fgfr. A set of Fgfr-targeted translation- and splice-blocking antisense morpholino oligonucleotides (MOs; Gene Tools LLC) had previously been designed for use in the lab. MOs bind to the translation start site or an intron-exon junction, and provide a steric block to the initiation of translation or the proper splicing of the pre-mRNA, respectively (Summerton and Weller, 1997). By interfering with these processes, MOs can cause a reduction in the amount of functional protein that is produced (Nasevicius and Ekker, 2000). The sequences of MOs I used are listed in Table 1.1. I focused on splice-blocking MOs because their efficacy can be examined via RT-PCR, and mRNAs transcribed *in vitro*, which require no splicing, can be used to rescue MO-injected embryos.

X. tropicalis or *X. laevis* embryos were injected with MO into both cells at the two-cell stage with the fluorescein standard control MO as a tracer, and fixed at stage 11 equivalent as gauged by uninjected controls. I assayed the effect of splice-blocking MOs on their target transcripts by RT-PCR (Figures 1.2–1.4), and I assayed the effect of each MO on mesoderm specification via *in situ* hybridization for the mesoderm marker *t* (Figures 1.2–1.4). *t* staining phenotypes of embryos were placed into four classes, as compared to uninjected controls: wildtype, mild reduction in *t* (slight reduction in intensity of *t* stain or thickness of band), moderate (stronger reduction in intensity of *t* stain or patchy staining), severe (*t* stain is very faint or completely absent). Examples of these phenotypes for *X. tropicalis* and *X. laevis* are given in Figure 1.2 and Figure 1.3, respectively.

The Fgfr1-targeting MO XtFGFR1(I3E4) targets the junction between intron 4 and exon 5 (exons 1–6 code for the signal sequence and first two Ig domains of the protein). In *X. tropicalis* embryos injected with this MO, RT-PCR showed a slight decrease in *fgfr1* transcript in injected embryos. A larger band representing intron inclusion was not detected, but this could be due to the large size of the intron (3 kb), so an intron-specific primer may be required to detect intron inclusion. By *in situ* hybridization for *t*, MO-injected embryos showed a dose-dependent reduction in mesoderm at doses of 10 ng or greater (Figure 1.2). Embryos injected with 10 or 20 ng of MO and allowed to develop past stage 11 exhibited gastrulation defects. This result is consistent with a failure to properly specify mesoderm because gastrulation is driven in part by involution of mesodermal cells.

Figure 1.3 shows the effects of the Fgfr3-targeting MO XtFGFR3(I3E4). RT-PCR with primers in the exons flanking the MO binding site showed a slight increase in transcript in injected embryos, consistent with the MO stabilizing the *fgfr3* transcript. As with the Fgfr1 MO, an intron-specific primer would be required to test whether this MO causes intron inclusion. Injected embryos exhibited a dose-dependent reduction in *t* expression: a mild phenotype was seen with as little as 20 ng total injected, but the most dramatic loss of *t* was achieved with the relatively high dose of 80 ng. These embryos also later exhibit the gastrulation defects expected from embryos with improper mesoderm specification. *X. laevis*

embryos injected with the Fgfr3-targeting XLFR3-1 MO, a translation-blocking MO, exhibited wildtype *t* expression (data not shown), but this result could be due to a failure of the MO to sufficiently reduce Fgfr3 protein.

Injection of the Fgfr4-targeting MO MOSAFR4_I3E4 into *X. laevis* occasionally, but not consistently, caused a reduction in *t* staining. Since this MO caused gastrulation defects even when mesoderm specification was unaffected (data not shown), I concluded that it might be toxic. The Fgfr4 MO MOSAFR4_I2E3 caused a dose-dependent reduction in mesoderm specification in *X. laevis* (Figure 1.4). Embryos injected with as few as 40 ng of MO exhibited inclusion of the targeted intron, but a strong knockdown of mesoderm specification was not achieved until the very high dose of 160 ng was injected.

The Fgfr2-targeting MO XtFGFR2(I4E5) did not have a repeatable phenotype in *X. laevis*, and had no clear phenotype in *X. tropicalis* (data not shown). This result is consistent with Fgfr2 being dispensable for mesoderm specification, but it is also consistent with the MO failing to cause a reduction in Fgfr2 protein. Without further investigation we cannot distinguish between these possibilities.

The knockdown experiments described above suggest that Fgfrs 1, 3, and 4 are each required for proper mesoderm specification, however, rescue experiments using mRNAs for the targeted gene were needed to prove that the MO-associated phenotypes were due to knockdown of the targeted genes and not off-target effects or toxicity.

MO name	Sequence
XtFGFR1(I3E4)	5' CAAAATGGACCTGCAGAAAAGAAGG 3'
XtFGFR2(I4E5)	5' AGCTCCTAGCGTTAGGGATCAAACA 3'
XtFGFR3(I3E4)	5' ATGTCGGAGCTACAATAAAGAAAAC 3'
MOSAFR4_I2E3	5' CAGATGCCAACGAGTCTACGAGAAA 3'
MOSAFR4_I3E4	5' GGTCCAGTATGGTGCTGAATGATAG 3'
XLFR3_1	5' CAGAACAGGGTGACCATTCCCATAT 3'
Fluoresceinated standard control	5' CCTCTTACCTCAGTTACAATTTATA 3'

Table 1.1. Sequences of MOs used in this study.

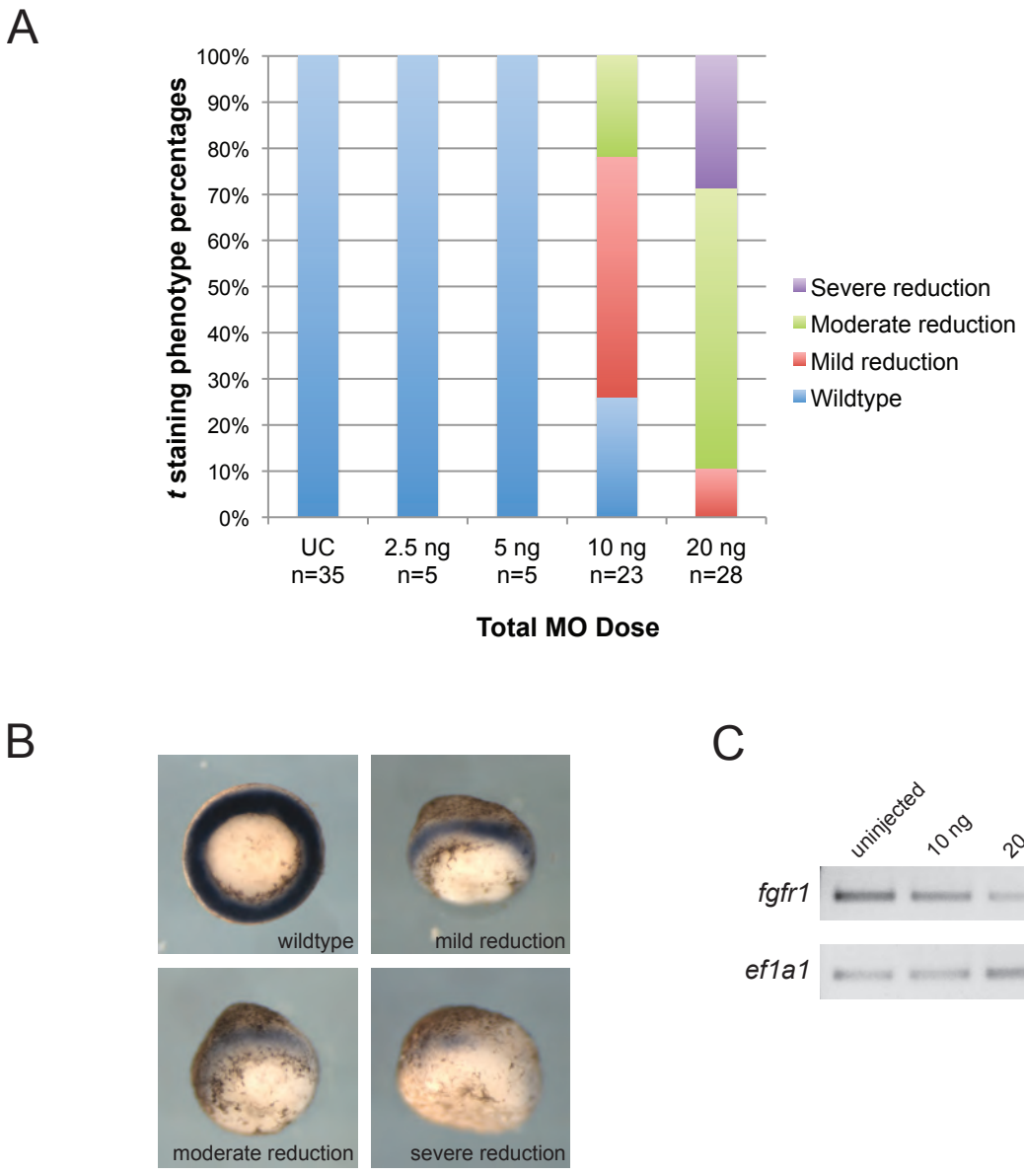


Figure 1.2. An MO targeting Fgfr1 causes a dose-dependent reduction in mesoderm specification. *X. tropicalis* embryos were injected into both cells at the two-cell stage with the *fgfr1* splice-blocking MO XtFGFR1(I3E4), and fixed at stage 11. The total dose of MO is given. (A) *t* staining phenotypes resulting from MO injection, as gauged by *in situ* hybridization for the *X. tropicalis t* transcript. UC, uninjected control embryos. (B) Examples of *t* phenotypes tallied in A. Wildtype, vegetal view; others vegetal-lateral view. Embryos injected with 10 or 20 ng of MO do not successfully gastrulate: failure to properly form a blastopore is evident in the "severe reduction" image. (C) RT-PCR for *fgfr1* shows a mild decrease in transcript in MO-injected embryos. *fgfr1* primers lie in exon 4 and exon 5 (the MO is complementary to the intron 4/exon 5 junction). *ef1a1* is a loading control. The RT– cDNA reaction contained RNA isolated in parallel with the others, but did not contain reverse transcriptase.

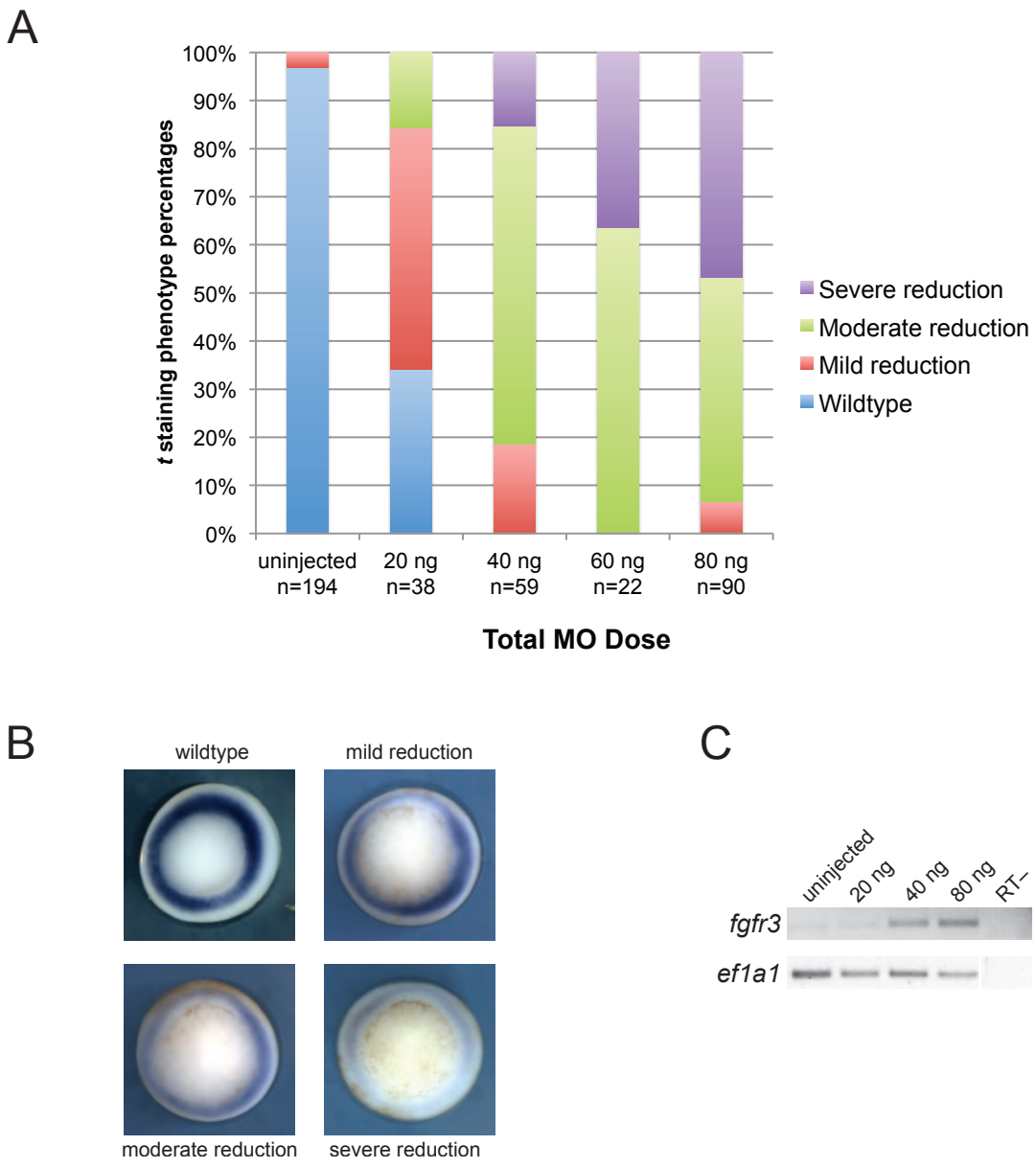


Figure 1.3. An MO targeting Fgfr3 causes a dose-dependent reduction in mesoderm specification. *X. laevis* embryos were injected into both cells at the two-cell stage with the *fgfr3* splice-blocking MO XtFGFR3(I3E4), and fixed at stage 11. The total dose of MO is given. (A) *t* staining phenotypes resulting from MO injection, based on *in situ* hybridization for the *X. laevis t* transcript. UC, uninjected control embryos. (B) Examples of *t* phenotypes tallied in A. Vegetal views. (C) RT-PCR for *fgfr3* shows an increase in transcript in MO-injected embryos. *fgfr3* primers lie in exon 4 and exon 5 (the MO is complementary to the intron 4/exon 5 junction). *ef1a1* is a loading control. The RT– cDNA reaction contained RNA isolated in parallel with the others, but did not contain reverse transcriptase.

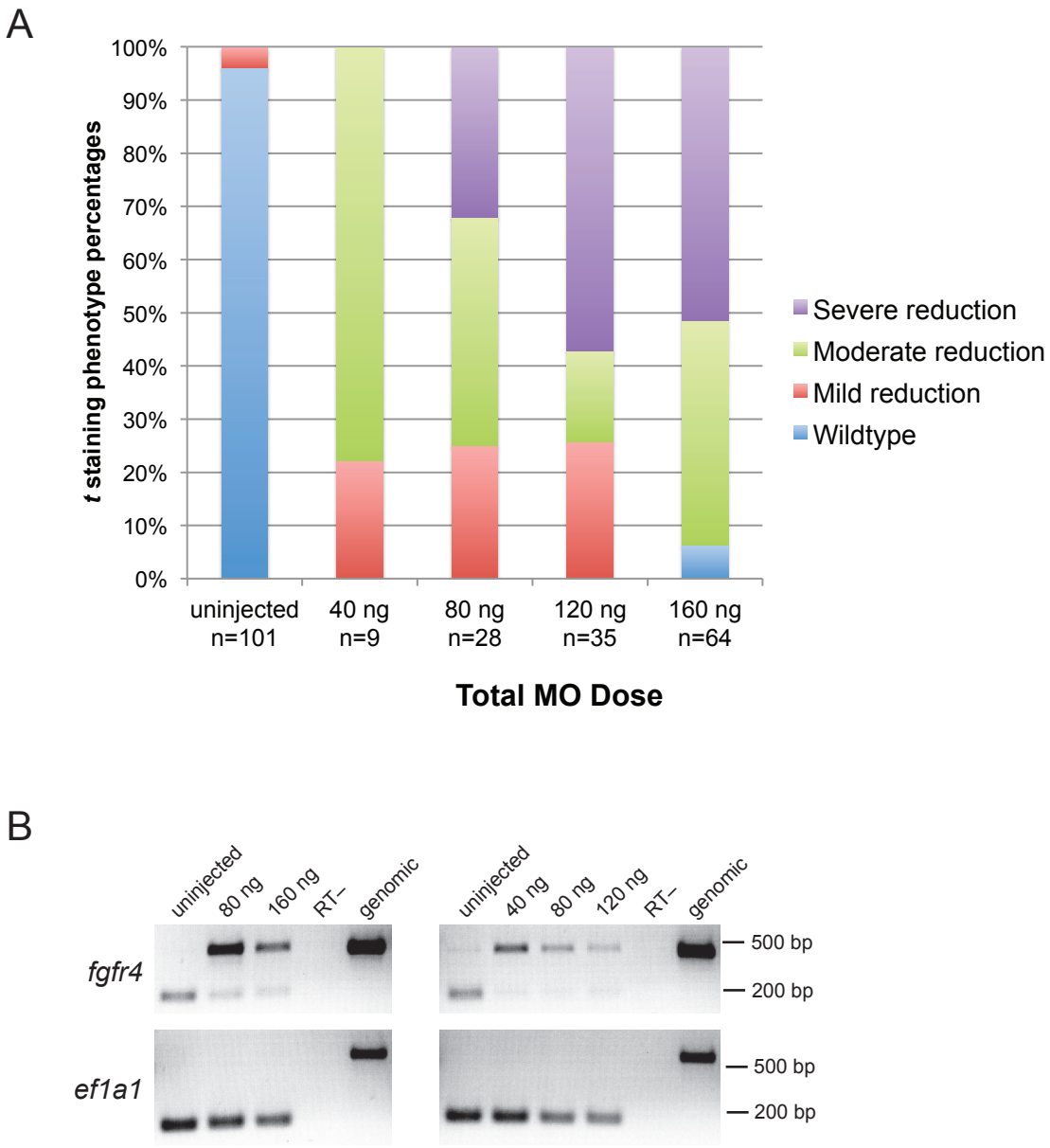


Figure 1.4. An MO targeting Fgfr4 causes a dose-dependent reduction in mesoderm specification. Embryos were injected with the *fgfr4* splice-blocking MO MOSAFR4_I2E3 into both cells at the two-cell stage and fixed at stage 11. The total dose of MO is given. (A) *t* phenotypes resulting from MO injection into *X. laevis*. *in situ* hybridization was performed using a probe for the *X. laevis t* transcript and scored as in Figure 1.3. UC, uninjected control embryos. (B) RT-PCR for *fgfr4* shows intron inclusion and a decrease in properly spliced transcript in embryos injected with as little as 40 ng of MO. *fgfr4* primers are in exon 3 and exon 4 (the MO is complementary to the intron 3/exon 4 junction). *ef1a1* is a loading control. The RT– cDNA reaction contained RNA isolated in parallel with the others, but did not contain reverse transcriptase. The band in the genomic lane was amplified from genomic DNA and thus is the size expected in the case of intron inclusion.

IIb. Fgfr overexpression

In preparation for MO rescue experiments, I tested the effects of overexpressing *fgfr* genes via injection of mRNA. I initially injected mRNAs transcribed from plasmids made by Russell Fletcher, which were derived from ESTs (Appendix). LacZ mRNA was used as a tracer. Embryos were fixed at stage 11, stained for LacZ, and then stained for *t* expression via *in situ* hybridization.

I expected that overexpression of Fgfrs would cause an expansion of mesoderm, because Fgf ligand is sufficient to induce mesoderm in animal caps, as is Ras, which transduces Fgf signal through the MAPK and PI3K pathways (Isaacs et al., 1992; Umbhauer et al., 1995; Song and Slack, 1996; Lombardo et al., 1998; Carballada et al., 2001; Fletcher et al., 2006). I was surprised to find that Fgfr overexpression did not have very much effect on *t* expression (Figure 1.5). Overexpression of *X. tropicalis fgfr1* or *fgfr3*, or *X. laevis fgfr4b* had no effect. Overexpression of *X. laevis fgfr2* did cause an expansion of *t*, but it was relatively mild compared to the effect of the mesoderm-inducing ligand *fgf8b*, 5 pg of which is sufficient to broadly expand the mesoderm toward the animal pole (Fletcher et al., 2006). Interestingly, we noticed that *X. laevis fgfr4a* actually caused a mild reduction in *t* staining where the LacZ tracer intersected the marginal zone of injected embryos. The mild effects on mesoderm specification resulting from *fgfr* overexpression raised concerns about whether the mRNAs may not be robustly translated because they contained some UTR sequence (given in Figure 1.5). I made UTR-free versions of the *fgfr3* and *fgfr4a* plasmids; injection of mRNAs transcribed from these plasmids causes similar results to their UTR-containing versions (data not shown). As a positive control, I tested the effect of a constitutively-dimerizing *Ciona intestinalis* Fgf receptor (*torso-CiFGFR*, Figure 1.5) (Shi and Levine, 2008). Injection of mRNA caused an expansion of *XBra* staining similar in appearance to that seen with *fgfr2*, but only in a minority of embryos. Umbhauer and colleagues showed that the constitutively-dimerizing torso-Fgfr4 is a poor inducer of mesoderm in animal caps (Umbhauer et al., 2000), and indeed, I found that at low doses, injection of this mRNA had no effect but at very high doses it reduced *t* expression (Figure 1.5). This result is consistent with the intracellular domain of Fgfr4 being a poor activator of MAPK, as previously reported (Wang et al., 1994).

The lack of a dramatic increase in mesoderm resulting from *fgfr* mRNA overexpression suggests that the ligand may be a limiting factor, as ligand is required for receptor dimerization and thus the transphosphorylation leading to signal transduction (Böttcher and Niehrs, 2005). This result is also consistent with mechanisms such as negative feedback preventing the induction of too much mesoderm at the expense of other tissues. For example, expression of the negative Fgf pathway regulator Spred is induced by Fgf signaling in frogs and inhibits mesoderm formation (Sivak et al. 2005).

Yamagishi & Okamoto showed that *fgfr1* knockdown in *Xenopus* anteriorizes the neural plate, but *fgfr4* knockdown posteriorizes it. Although ligand-independent constitutively active (ca-) *fgfr4* and *fgfr1* both posteriorize the neural plate, *ca-fgfr4* does so in a less potent manner (Yamagishi and Okamoto, 2010). The authors concluded that *fgfr4* competes with *fgfr1* for *fgf* ligand, but that since it transduces the signal less strongly, it serves to dampen the signal. This model is consistent with my result that *fgfr4* overexpression causes a

reduction of mesoderm specification in the marginal zone, as well as the induction of *t* expression more animally observed only at high doses of *fgfr4*.

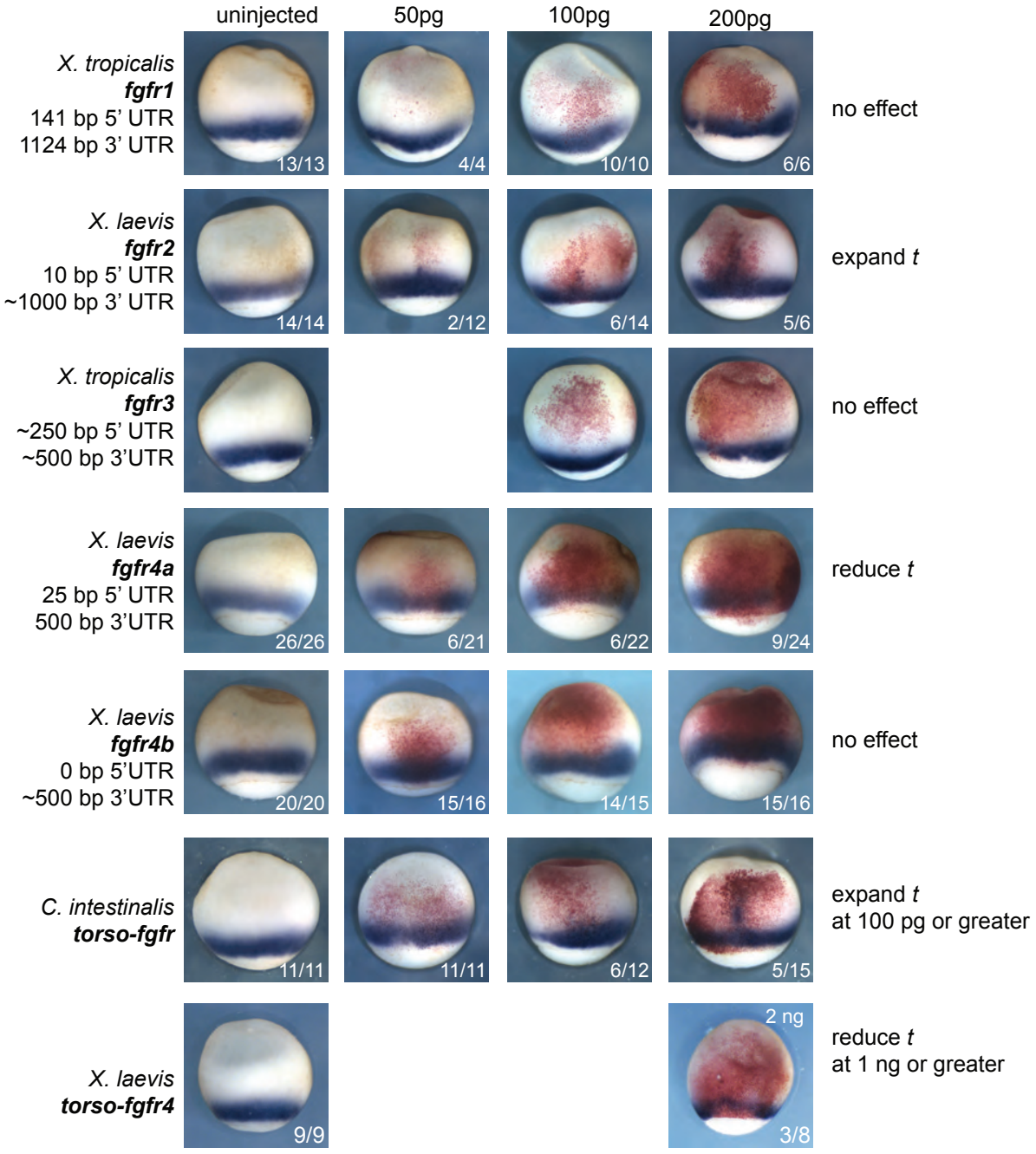


Figure 1.5. Effect of Fgfr overexpression in *X. laevis*. Lateral views of embryos stained for *t* by *in situ* hybridization. The text on the right describes the phenotypic trend for a given mRNA, and numbers tell how many assayed embryos had the phenotype depicted in each image. Where applicable, lengths of UTRs included in transcribed mRNAs are indicated.

IIc. Morpholino rescue experiments

Figure 1.6 outlines the strategy I used for MO rescue experiments. Since the extent of reduction in *t* staining resulting from MO injection varies from embryo to embryo (e.g. as shown in Figures 1.2 and 1.3), I chose to attempt unilateral rescue in embryos given equal doses of MO on both sides in order to provide an internal control for each embryo. Thus, whether or not the rescue was successful was gauged by comparing the MO + mRNA side of each embryo to the MO-only side of the same embryo.

I attempted rescue of the Fgfr3 and Fgfr4 MO knockdown phenotypes in *X. laevis* with as little as 50 pg and as much as 2 ng of the corresponding mRNA (with and without UTR sequence), and in combination with up to 50 pg of the mesoderm-inducing ligand *fgf8b* (Fletcher et al., 2006). Figure 1.7 shows some selected successful rescue attempts for the Fgfr3 MO. Notably, all successful rescue attempts for the Fgfr3 MO also included some *fgf8b* ligand. No rescue attempts with *fgfr3* mRNA alone were successful. As shown in Figure 1.8, the lower doses of *fgf8b* mRNA and/or *fgf8b* + *fgfr3* mRNAs expand the *t* stain. Higher doses cause the characteristic reduction of *t* expression in the marginal zone but expansion more animally: I interpret the reduction in the marginal zone as consistent with negative feedback upon high levels of Fgf signaling, but more animal regions don't normally receive the Fgf signal, and so they are induced to become mesoderm. Embryos displayed more *t* stain on the side with MO and mRNA than on the MO-only side. The requirement for Fgf ligand for the restoration of *t* expression to Fgfr3 MO-injected embryos confounds interpretation of whether the MO is indeed specific for the *fgfr3* transcript, but it does suggest at least that the MO specifically affects Fgf signaling.

None of my attempts at rescuing the Fgfr4 MO MOSAFR4_I2E3 using the dosage ranges listed above were successful. This could be because the phenotype is due to off-target effects or toxicity, especially given the large dose of MO required for a strong phenotype. Figure 1.8 shows a typical rescue attempt: there was never more *t* stain on the MO + mRNA side than the MO-only side of injected embryos, and it sometimes appeared that there was even less on the MO + mRNA side.

As with the Fgfr overexpression experiments, these experiments are consistent with very tight regulation of the expression and effects of genes in the Fgf pathway: it may be that rescue of Fgfr MO phenotypes is challenging or impossible because the members of this pathway must be in a very specific balance with one another for proper mesoderm specification and maintenance to occur.

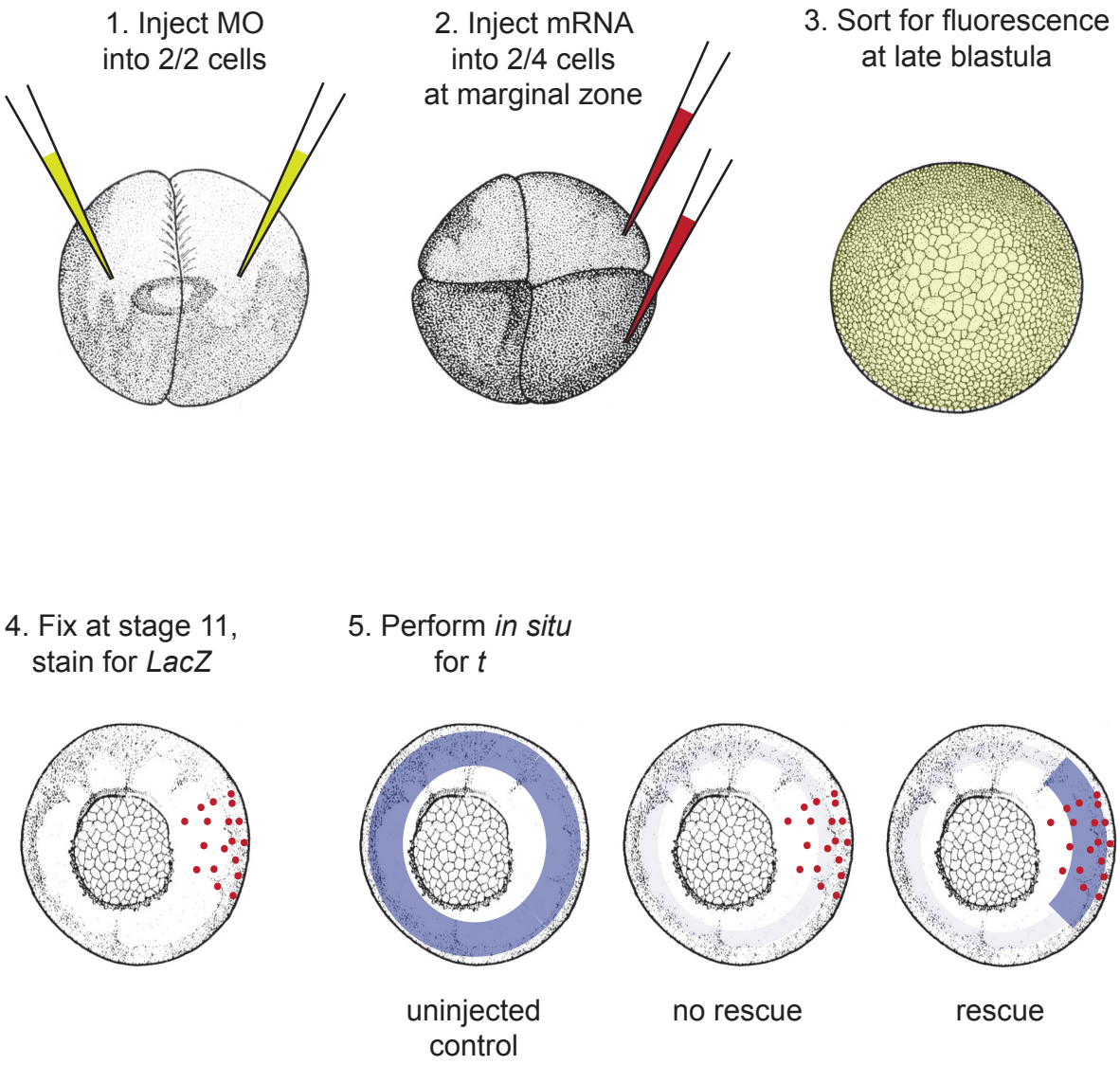


Figure 1.6. Scheme for MO rescue experiments. Adapted from (Nieuwkoop, 1994). (1–3) Animal view. (1) MO is injected into both sides of the embryo. (2) *fgfr* and/or *fgf8b* mRNA is injected into only one side of the embryo, so that the non-mRNA-injected side can serve as a comparison. (3) Fluorescein detection indicates successful MO injection. (4–5) Vegetal view. (4) *LacZ* stain (red) indicates successful mRNA injection. (5) If rescue is successful, there will be a greater amount of *t* expression (purple) in the marginal zone in mRNA-injected regions.

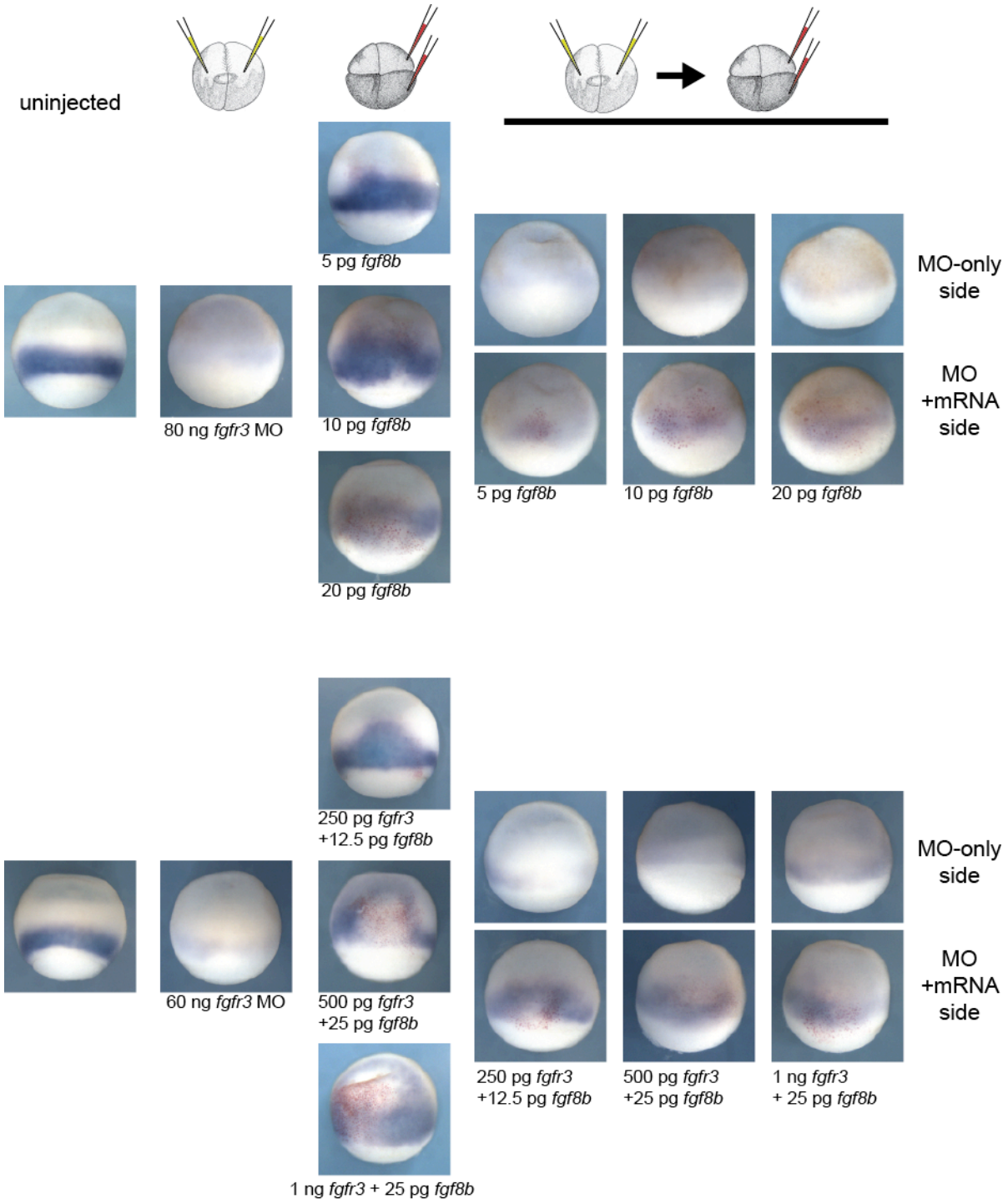


Figure 1.7. Selected Fgfr3 MO rescue experiments. Lateral views of embryos stained for *t* via *in situ* hybridization. Two separate experiments are shown. Rescue of Fgfr3 MO-injected embryos was only successful in conjunction with *fgf8b* mRNA.

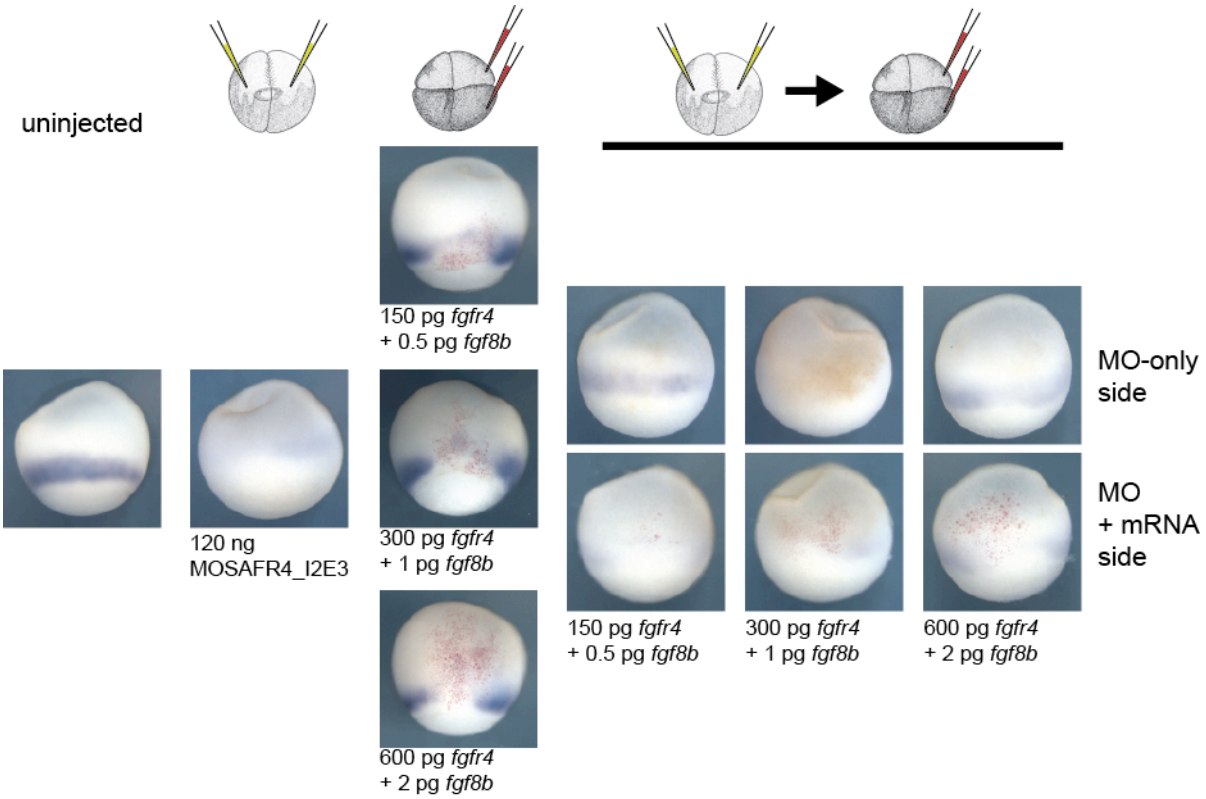


Figure 1.8. Fgfr4 MO-injected embryos could not be rescued. Lateral views of embryos stained for *t* via *in situ* hybridization. I was unable to rescue this MO phenotype, with *fgfr4* mRNA, *fgf8b* mRNA, or with *fgfr4* and *fgf8b* mRNAs together.

IId. Identity of marginal zone cells lacking *t* expression upon Fgf pathway manipulation

I wondered what the identity might be of marginal zone cells that had experienced a loss in *t* expression (*t*– cells) due to Fgfr3 MO injection, Fgfr4 overexpression, or Fgf8b overexpression. I performed a molecular expression analysis on embryos injected with these reagents, staining for markers of various tissues. Embryos were injected unilaterally so that in each embryo the injected side could be compared to the wildtype pattern on the uninjected side. I used the endoderm markers *bix4* and *sox17b* to test whether the endoderm was expanded into the marginal zone in treated embryos. In order to investigate whether the dorsal/ventral identity of the mesoderm was altered, I tested the paraxial mesoderm marker *myf5*, the ventral mesoderm marker *vent2*, and the axial mesoderm markers *chordin* and *goosecoid*. Results of this experiment are summarized in Table 1.2, and selected images for markers that showed a change are in Figure 1.9.

Embryos injected with 40 ng of the Fgfr3 MO had reduced expression of *myf5* and *chordin* on the injected side, and interestingly, most of these lost *chordin* expression bilaterally (Figure 1.9A). These results are consistent with a general loss of mesoderm, but *vent2*, the ventral mesoderm marker, appeared unaffected, as did *fgf8b* and *goosecoid*, suggesting that the expression of these genes may not require Fgfr3 and/or that their expression does not require high doses of Fgf signal (Table 1.2). None of the markers examined were expanded, so it remains unclear what identity the Fgfr3 knockdown marginal zone cells may have.

Embryos injected with the high dose of 1 ng of *fgf4* mRNA exhibited the characteristic loss of *t* in the mesoderm but expansion animally, and *fgf8b* expression showed a similar phenotype (Figure 1.9B). Although most injected embryos displayed an expansion of *sox17b* expression (Figure 1.9B), it was a mild effect, and the other endoderm marker, *bix4*, was mostly unaffected (Table 1.2). *vent2*, *goosecoid*, and *chordin* displayed wildtype expression patterns.

The effect of high doses of *fgf8b* on *t* expression was shown in Figure 1.7—it is characterized by a loss of *t* staining in the marginal zone but an expansion into the animal hemisphere, consistent with the triggering of negative feedback in the marginal zone but induction of mesoderm in more animal cells. In embryos injected with 30 pg of *fgf8b*, the expression of *goosecoid* and *chordin* is dampened on the injected side, but also appear to be expanded (Figure 1.9C). This could be due to an animal expansion of mesoderm, but it could also be due to a morphogenesis defect, so it remains unclear what specific effect *fgf8b* may have on dorsal mesoderm.

All in all, it remains unclear what identity *t*– marginal zone cells may have, though the data suggest that manipulation of the Fgf pathway is not sufficient for endodermal fate. Also, depending on the reagent used, not all mesoderm is affected by all of the treatments; for example, *goosecoid* is unaffected by Fgfr3 knockdown or *fgfr4* overexpression, but it is dampened by high doses of *fgf8b*. Fletcher and Harland (2008) showed that Fgf signaling is dispensable in the establishment of organizer fates but is required for their maintenance; perhaps the amount of the reduction in Fgf signaling resulting from Fgfr3 knockdown or

fgfr4 overexpression is not sufficient to prevent *goosecoid* expression, but the negative feedback resulting from *fgf8b* overexpression is sufficient to hamper it.

Treatment	Marker	Tissue	No clear effect	Expanded	Reduced
40 ng *fgfr3* MO	*t*	mesoderm			1 mild, 4 moderate, 7 severe
	myf5	paraxial mesoderm	1		7
	vent2	ventral mesoderm	5	1, dorsally	
	fgf8b	marginal zone	12		
	goosecoid	dorsal mesoderm	12		
	chordin	dorsal mesoderm	5		9
	bix4	endoderm	12		
	sox17b	endoderm	9		2, very slight
1 ng *fgfr4* RNA	*t*	mesoderm		2	5 reduced, 4 reduced in MZ and expanded animally
	vent2	ventral mesoderm	8	1	3
	fgf8b	marginal zone	1		11 reduced in MZ and expanded animally
	goosecoid	dorsal mesoderm	9	2, slight	1
	chordin	dorsal mesoderm	8	3, slight	
	bix4	endoderm	9	2	
	sox17b	endoderm	2	7	1
30 pg *fgf8b* RNA	*t*	mesoderm			13 reduced in MZ and expanded animally
	goosecoid	dorsal mesoderm			1 reduced, 4 reduced in MZ and expanded animally
	chordin	dorsal mesoderm	6		5, dampened and expanded

Table 1.2. Effect of Fgf pathway manipulation on selected markers. "MZ" refers to the marginal zone.

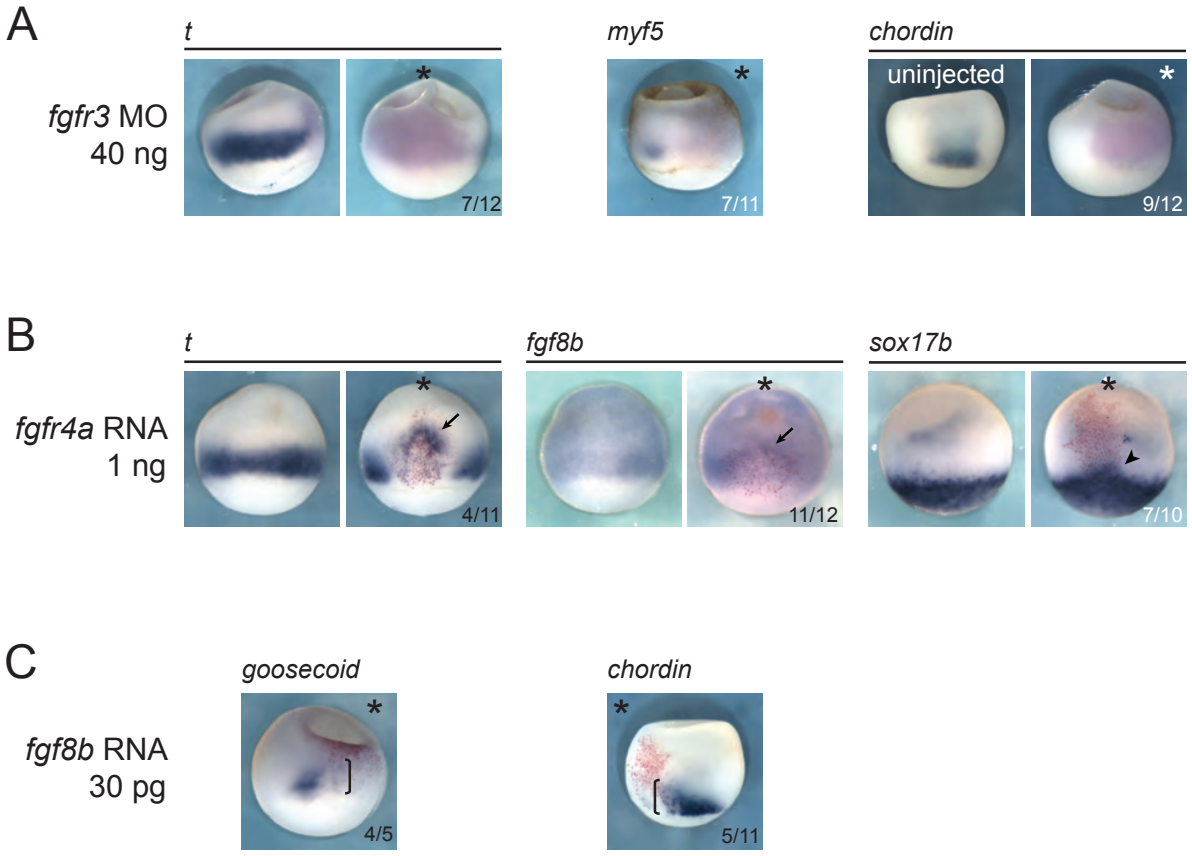


Figure 1.9. Effect of *fgfr3* knockdown, *fgfr4* overexpression, or *fgf8b* overexpression on mesendodermal markers. The injected side of each embryo is indicated with an asterisk. (A) Cells exposed to the Fgfr3 MO, which causes a dramatic reduction in *t* staining, also fail to express the paraxial mesoderm marker *myf5* and the dorsal mesoderm marker *chordin*. Pink staining marks the location of the fluoresceinated control MO, which was used as a tracer. (B) Overexpression of high amounts of *fgfr4* has a similar effect on *t* and the ligand *fgf8b*, which are reduced in the marginal zone but expanded animally (arrows). The endoderm marker *sox17b* expanded slightly into the marginal zone (arrowhead). Red stain marks the location of *LacZ*, the tracer used for the mRNA. (C) The dorsal mesoderm markers *goosecoid* and *chordin* are dampened but expanded slightly on the side of embryos injected with a high dose of *fgf8b* (braces). *t*, *fgf8b*, and *sox17b* images, lateral view; others, dorso-lateral view.

III. Alternative splicing of *X. tropicalis* Fgfrs in the extracellular Ig domain D3

As mentioned above, it is known that in mammals the transcripts of *fgfr*s 1–3 exhibit alternative splicing such that the C-terminal half of D3, the most membrane-proximal Ig domain, is encoded by mutually exclusive exons (Eswarakumar et al., 2005) (Figure 1.10A), resulting in isoforms that thus may differ in ligand-binding specificity (Miki et al., 1992; Chellaiah et al., 1994). Expression patterns of the *X. laevis* Fgfrs have been published (Friesel and Brown, 1992; Golub et al., 2000; Pope et al., 2010), but none of these studies parsed the expression of alternative spliceforms. Robert Lea and colleagues reported the temporal and spatial expression patterns for *X. tropicalis* Fgf ligands and receptors (Lea et al., 2009). Although they used RT-PCR to investigate the alternative splicing that affects whether D1 is included in *fgfr1* and *fgfr2* transcripts, they did not investigate the alternative splicing of the *fgfr*s in D3.

Therefore, I sought to characterize the expression of the *X. tropicalis* Fgfr D3 alternative spliceforms. Alternative splicing of the *fgfr*s in D3 has not been reported in *Xenopus*, and *X. tropicalis* genome assembly annotations give only one form (xenbase.org).

In order to find out where in the *fgfr* genes to look for alternative D3 exons, I first used the annotated *X. tropicalis* gene model for each receptor (assembly version 4.1) (Hellsten et al., 2010) and the protein domain-prediction website SMART (http://smart.embl-heidelberg.de/) to determine which exons coded for which parts of the protein. *fgfr2* had three ESTs that appeared to contain a different exon 5' to the exon in the gene model that codes for the C-terminal half of D3 (metazome.org). I developed an approximation of the sequence of the non-annotated exon by comparing these EST sequences to the reference, and assessing what sequence could successfully form an Ig domain in conjunction with the IIIa exon. I confirmed the sequence of this IIIb exon by cloning it from cDNA using primers flanking the alternatively spliced region, and sequencing several clones (because the flanking primers would also amplify the exon in the gene model). Thus, I found that the C-terminal D3 exon included in the *fgfr2* gene model is the IIIc form, but this gene also has a IIIb exon (Figure 1.10).

I followed similar procedures to identify alternatively spliced exons in *fgfr1* and *fgfr3*. *fgfr1* had one EST with a partial alternative D3 C-terminal exon (metazome.org), and I was able to approximate the full length of the exon using BLAST with the *fgfr2* IIIb exon against the region of the genomic sequence in *fgfr1* in which it lies. By cloning and sequencing as above, I confirmed that the *fgfr1* gene model contains the IIIc exon but this gene also has a IIIb exon that can form the C-terminal half of D3. *fgfr3* has no EST coverage in the region of the gene that codes for D3 (metazome.org), but I was able to identify the missing *fgfr3* IIIb exon by BLASTing against the *fgfr2* alternatively spliced exons and using SMART to confirm that it could form an Ig domain with the IIIa exon (Figure 1.10).

I have BLASTed various alternatively spliced IIIb and IIIc D3 exons against the genome sequence around the *fgfr4* D3 exons, but I have not found any evidence in the genome for a second exon that could form the C-terminal half of D3. As mammalian *fgfr4*s are also not alternatively spliced in D3 (Itoh and Ornitz, 2004; Eswarakumar et al., 2005), this suggests

that the second isoform of this gene was lost before the divergence of the lineages leading to frogs and mammals. The sea urchin *Strongylocentrotus purpuratus* has one Fgfr gene that is alternatively spliced in the same manner as vertebrate Fgfrs, so that of the last common ancestor of all deuterostomes must have been as well (Itoh and Ornitz, 2004). The sequences for the IIIb and IIIc exons of *X. tropicalis* Fgfrs 1–3 are listed in Table 1.3.

To assess whether and when each of these *X. tropicalis* alternative D3 spliceforms is expressed, I performed RT-PCR using spliceform-specific primer pairs (Figure 1.10B). Interestingly, for each receptor, only one spliceform is detected before the maternal to zygotic transition (stage 6.5; zygotic transcription begins at stage 8). For *fgfr1*, the IIIc form is maternal, but for *fgfr*s *2* and *3*, the IIIb form is maternal. The steep drop in expression level of the *fgfr3* IIIb form after stage 10.5 suggests that there is a mechanism in place for degrading the maternal transcript to ensure that levels of this isoform are low during neurulation.

In order to evaluate the spatial expression patterns of *fgfr* D3 isoforms in *X. tropicalis*, I cloned and transcribed *in situ* probes from the alternatively spliced exons. I also made probes against the 3'UTRs of these genes, which should detect all Ig domain isoforms (Appendix). *in situ* results for *fgfr1* and *fgfr2* are shown in Figure 1.11. Some notable differences between the spliceforms are evident for both genes. The tailbud, kidney, and midbrain-hindbrain boundary expression of *fgfr1* are only detected by the probe for the IIIc form, but only the IIIb form is detected in the otic vesicle. Staining for *fgfr2* was not as robust as that for *fgfr1*, but nonetheless, the staining suggests that the midbrain-hindbrain boundary expression is IIIc-specific, whereas the otic vesicle expression is specific to the IIIb form.

*in situ*s for *fgfr3* alternatively spliced D3 exons stain only faintly and in a manner that resembles sense probe controls, even when a more sensitive protocol (omitting the RNAse step and the most stringent SSC washes) is used. This may be simply due to the fact that these alternatively spliced exons, and thus the probes against them, are only 145-148 bp long, which is quite short for *Xenopus in situ* probes. Locked nucleic acid probes may be needed for successful *in situ* hybridization for these isoforms (Darnell et al., 2010).

The differences in expression patterns of D3 isoforms of the *Xenopus* Fgf receptors raise questions about the roles played by these isoforms, as well as how their expression is regulated. For example, perhaps the maternally deposited forms are required for mesoderm specification and the zygotic-only forms play roles in morphogenesis: one might imagine that by ensuring that the former are present earlier, mesoderm specification is completed before gastrulation occurs. As D3 plays a role in ligand specificity, it also seems likely that the regulation of the differential splicing of Fgfrs in a given cell is important for controlling which Fgf signals are transduced into that cell. So far, vertebrate Fgf-Fgfr specificity has only been examined *in vitro*, using cells that do not normally express Fgfs or Fgfrs (Ornitz et al., 1996; Zhang et al., 2006), but this work suggests that Fgf ligand-receptor interactions are tightly regulated during development. Furthermore, given that the different isoforms of Fgfrs 1–3 may have different roles and activities, it will be important to take this alternative

splicing into account when investigating the roles played by Fgf receptors during development.

Taken together, my *fgfr* MO and mRNA overexpression experiments are consistent with the model that although all four receptors play roles in mesoderm specification, Fgfr4 serves to dampen the transduction of the Fgf signal and thus restrict mesoderm specification to the marginal zone. My characterization of Fgfr 1–3 alternative spliceforms suggests that the precise regulation of their expression plays a role in controlling where and when Fgf signals are transduced during development.

Exon	Sequence
Fgfr1 IIIb	CACTCTGGGATTAATAGCTCAGACGCAGAAGTTTTGACCCTTTACAATGTGA CTGAGGCAGAGAGTGGGGAATACATATGTAAAGTGTCCAATTATATTGGTGA GGCCAATCAGTCTGCGTGGCTTACCGTCACCAGACCCGTGACAAAAG
Fgfr1 IIIc	ACTGCAGGAGTCAACACCTCGGACAAGGATATGGAGGTTCTCCACCTGAG AAATGTTACTTTTGAGGATGCTGGCCAGTATACCTGCTTGGCCGCTAACTCC ATTGGGATATCTCATCATTCTGCATGGTTGACCGTTCTTGAAG
Fgfr2 IIIb	CGCTCAGGAATTAACAGTTCCAGTGCTGAAGTGCTGAAACTGTACAATGTG ACAGAAGAGGACGCAGGGGAATATATATGTGCGGTCTCCAATTATATAGGAG AGGCCAACAAGTCTGCCTGGCTCACGGTGGAGCGTGAAAAAG
Fgfr2 IIIc	GCGGCTGGAGTTAACGTTACGGACGAAGAGATAGAAGTCTTGTATGTCAGG AATGTTTCTTTTGAGGATGCTGGGGAATATACTTGTATAGCTGGAAATTCTAT TGGGATTTCTCAACATTCTGCCTGGTTGACGGTTCATCCAG
Fgfr3 IIIb	TCTTTCACCAATGGCACTGAAGTCGATACTACCCTAAGTCTAAAAAATGTGA CAGAGACCCATGAAGGACAGTATGTGTGTAGAGCCAACAATTTCATAGGCG TAGCTGAGGCGTCCTTTTGGCTCCACATTTACAAACCAGCATCAG
Fgfr3 IIIc	CCTACTGGTGTTTACTCTTCGGATAAGGATCTTGAGGTGCTGGTTTTACGCA ATGTGTCCTTTGAGGATGCTGGGGAATATACTTGTCTGGCTGGGAATTCTAT TGGCTATTCACATCACACTGCTTGGCTGACGGTTCTCCCAG

Table 1.3. Sequences of *X. tropicalis* Fgfr D3 alternatively spliced exons.

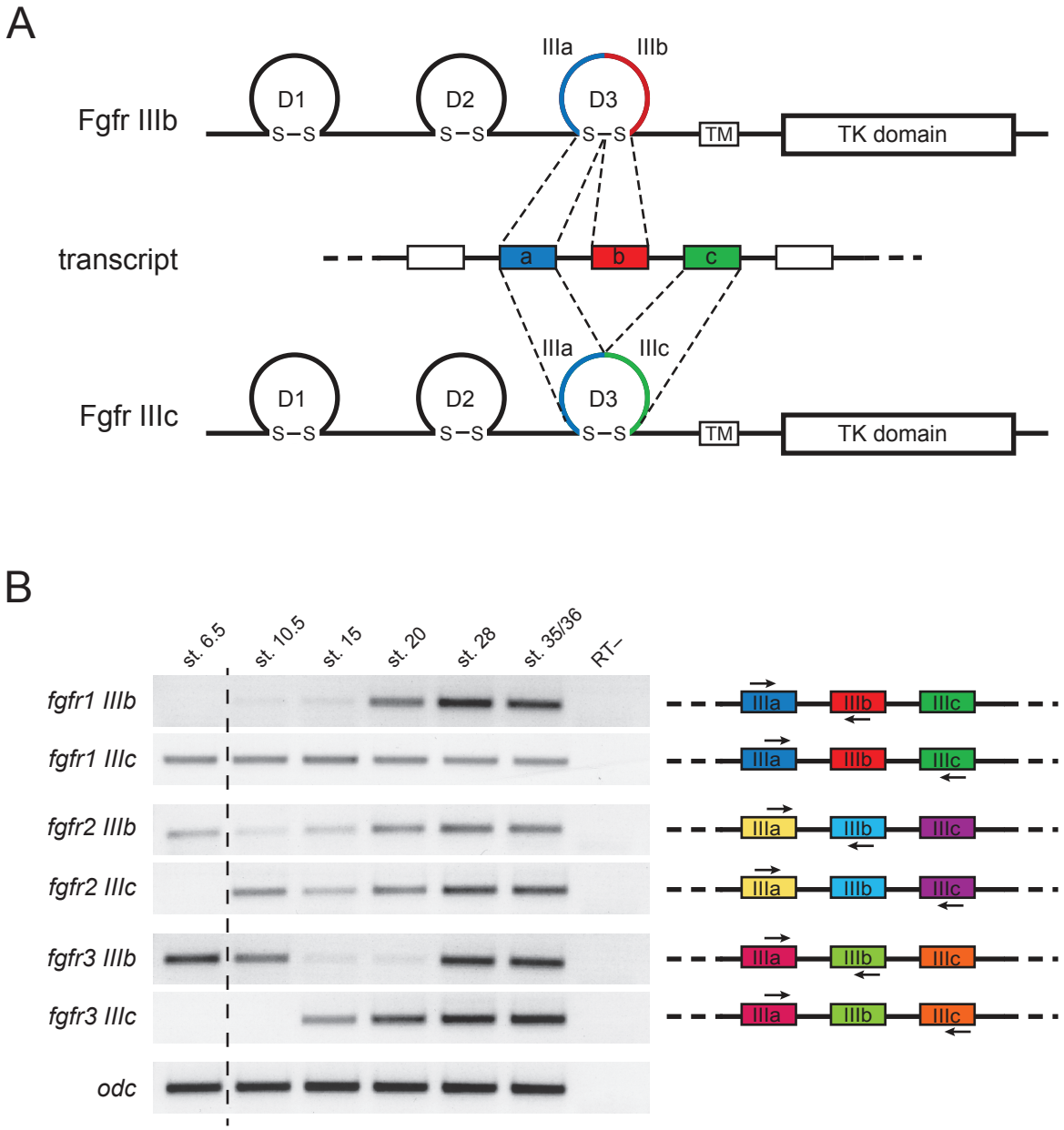


Figure 1.10. Alternative splicing of *X. tropicalis fgfr* transcripts. (A) Adapted from (Eswarakumar et al., 2005). Fgfrs have three extracellular Ig domains numbered D1–D3, a transmembrane domain, and an intracellular tyrosine kinase domain. Alternative splicing, mediated by the use of mutually exclusive exons coding for the C-terminal half of D3 and resulting in forms called IIIb and IIIc, is thought to affect ligand binding. (B) RT-PCR for *fgfr* D3 isoforms. For each gene, only one transcript is detected before the maternal to zygotic transition of gene transcription (dotted line). *odc* is a loading control.

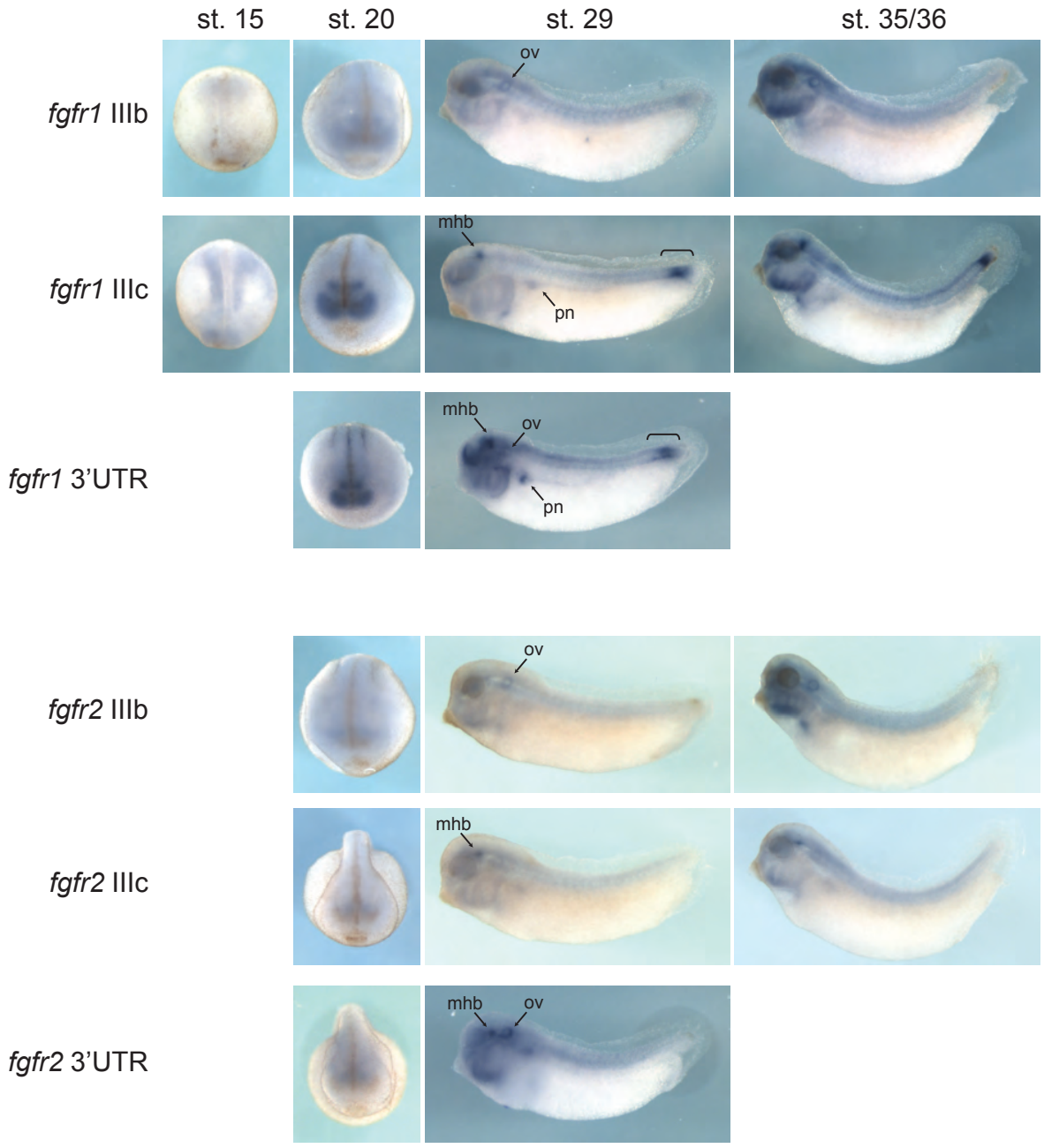


Figure 1.11. Expression of Fgfr1 and Fgfr2 D3 alternative spliceforms in *X. tropicalis*. mhb, midbrain-hindbrain boundary; ov, otic vesicle; pn, pronephros. Taibud expression of *fgfr1* is marked by a brace.

IV. Materials and Methods

Embryo culture

X. laevis embryos were obtained and cultured as described in (Sive, 2000). *X. tropicalis* embryos were obtained via *in vitro* fertilization as described in (Khokha et al., 2002), or via natural mating: for overnight natural matings, males were boosted with 100 units (U) human chorionic gonadotropin (HCG; Intervet) and females were boosted with 200 U HCG. For daytime natural matings, females were primed the night before with 10 U HCG, and the following morning males were boosted with 100 U HCG and females were boosted with 200 U HCG. *X. tropicalis* embryos were dejellied and cultured as described in (Khokha et al., 2002).

MO injection

MOs were injected into two-cell stage *X. tropicalis* or *X. laevis* embryos, with the Gene Tools fluorescein-conjugated standard control oligo as a tracer (Table 1.1). *X. tropicalis* injection volumes ranged from 0.5–2 nl/cell, and *X. laevis* injection volumes ranged from 2.5–10 nl/cell.

For some unilateral MO injections (Figure 1.9), an anti-fluorescein antibody was used to trace the fluoresceinated control MO. After *in situ* hybridization was performed, embryos were washed in maleic acid buffer (MAB), blocked for one hour in 2% BMB in MAB, and incubated for 4 hrs at room temperature with the alkaline phosphatase-conjugated anti-fluorescein antibody (Roche) diluted 1:3000. After being washed in MAB, stain was developed with Magenta Phos in alkaline phosphatase buffer with levamisol, then refixed with Bouin's fix.

in situ hybridization

in situ hybridization was performed essentially as described in (Sive, 2000), with the alkaline phosphatase substrate BM purple (Roche). The prehybridization incubation was shortened to as little as one hour for *X. tropicalis* embryos. For each probe on a given set of injected embryos, all color reactions were developed for the same amount of time. Antisense probes were transcribed using Sp6, T7, or T3 enzyme. The plasmids used to make them are listed in the Appendix.

mRNA injection

mRNA was transcribed from restriction-digested plasmids (Appendix) using the mMessage kit (Ambion). mRNAs were injected into one or two cells at the two-cell stage, or into two cells at the four-cell stage, with *LacZ* mRNA as a tracer. Occasionally, *katushka* mRNA (Scherbo et al 2007) was used as a tracer for sorting mRNA-injected embryos, but more frequently *LacZ* staining was used.

LacZ expression was assayed in partially fixed (30 min in MEMFA) embryos via incubation with Red Gal (Research Organics) at 4°C, room temperature, or 37°C. *LacZ* stain was developed and embryos were fixed further in MEMFA prior to *in situ* hybridization.

RT-PCR

RNA was isolated from pools of 3–10 embryos using Phenol:Chloroform extraction or TRIzol reagent (Invitrogen). cDNA was transcribed using M-MLV Reverse Transcriptase (Promega) and random hexamers. RT– cDNA reactions were performed with a subset of the RNAs isolated in a given batch, but no reverse transcriptase.

Where possible, genomic DNA from wildtype embryos of the appropriate species was used as a positive control and as a measure of intron-containing amplicon size.

PCR conditions: 30 cycles, 58°C or 60°C annealing, with up to 2.5 min extension time. RT–PCR primers are listed in the Appendix.

Chapter Two: Genotyping by sequencing in *X. tropicalis* and *X. laevis*

I. Introduction

X. tropicalis genome assembly version 4 (Hellsten et al., 2010) was constructed from paired Sanger sequence reads from libraries of various insert sizes, which included plasmid libraries of 2.95 kb and 8.30 kb average insert size, and a fosmid library of 38.5 kb insert size. Two *X. tropicalis* BAC libraries were employed, one from a male (CHORI, average insert size 175 kb) and one from a female (ISB-1, average insert size 57.4 kb). The total estimated sequence depth was 7.68X (Hellsten et al., 2010). The sequencing of paired ends of clones with large inserts of known size is particularly useful for the bridging of gaps. One key finding detailed in Hellsten et al. (2010) is the conservation of large syntenic regions between tetrapods, specifically between the frog, chicken, and human genomes. The main drawback to the version 4 assembly was its fragmentation: only half of the assembled sequence was in scaffolds greater than 1.56 Mb. In particular, this fragmentation presented hurdles for genetic mapping of mutations, although they could be overcome somewhat by utilizing synteny with chicken and/or human (Abu-Daya et al., 2009; Bhattacharya, 2010).

X. tropicalis assembly version 7 was constructed using the same sequence reads as version 4, but with improved assembly methods. Jerry Jenkins and Jeremy Schmutz first used the improved ARACHNE assembler from the Broad institute to assemble so-called supercontigs, which were then linked together using a microsatellite-based genetic map (Wells et al., 2011) and synteny data from chicken and human in order to construct chromosome-scale scaffolds. Neither of these data sets is completely reliable, however; in particular, the genetic map reported in Wells et al. displays a disturbing amount of intermixing of markers that lie on different scaffolds.

The advent of next-generation (nextgen) high-throughput sequencing, with its relatively low cost for high—and ever increasing—amounts of data, has fundamentally changed our approach to genome sequencing and analysis (Bentley, 2006; Metzker, 2010). The key innovation in this type of sequencing the use of reversible terminator nucleotides, which allow for the continued addition of exactly one nucleotide at a time onto a growing DNA strand (Bentley et al., 2008). In standard Illumina library preparation, DNA is fragmented by shearing, and adapter oligos are ligated onto the ends of these DNA fragments. The DNA is then amplified in a theoretically unbiased manner by PCR using primers specific to the adapters. Next, the DNA is denatured and applied to a "flow cell," the surface of which is bound to oligos complementary to the adapter/PCR primer sequences on one end or the other of each DNA strand (Figure 2.2). The DNA strands bind to the flow cell surface and are amplified in place to create clusters of DNA, each of which contains many copies of the same piece of DNA. Sequencing by synthesis is accomplished using fluorescent reversible terminator nucleotides, each base having a different color: one base is added at a time to each DNA fragment, and the clustering facilitates the robust detection of this process. Nextgen sequencing can generate billions of bases of sequence per sequencing run (Bentley et al., 2008).

Nextgen sequencing has made it feasible to use meiotic mapping of frequent single nucleotide polymorphism (SNP) markers for genetic map construction and to construct and verify chromosome-scale genome assemblies (Davey et al., 2011). The great power of these approaches is the ability to discover and genotype thousands of markers across the genome in a single sequencing run.

The first description of a reduced representation approach for the identification of SNP markers used shotgun Sanger sequencing. Altshuler and colleagues utilized a restriction digest to reproducibly sample the same markers in the human genome across individuals (Altshuler et al., 2000), and that approach has now been adapted for next-generation sequencing. The use of adapters incorporating a different barcode for each sample allows for the multiplexing of individuals into one sequencing run, which reduces the cost of sequencing per individual (Parameswaran et al., 2007; Craig et al., 2008). Multiplexing does result in reduced coverage per individual, but one need not resequence an entire genome in order to obtain a high number of usable markers, and the reduced representation approach allows for sufficient coverage at markers of interest. The choice of restriction enzymes, or a combination thereof, that cut more or less frequently can be used to control the density of marker coverage.

Baird and colleagues first demonstrated the so-called RAD tags (restriction-site associated DNA sequencing) approach to reduced-representation sequencing (Baird et al., 2008). By their method, DNA is fragmented via restriction enzyme, followed by ligation to a first set of adapters, then another round of fragmentation (by shearing, this time) and ligation of a second set of adapters before amplification. The Buckler lab at Cornell developed a streamlined version of reduced representation multiplex Illumina sequencing, featuring fewer fragmentation, ligation, and cleaning steps, and dispensing with size selection (Elshire et al., 2011). In the Elshire approach, which they call genotyping by sequencing (GBS), DNA and adapters are pipetted into wells of a 96-well plate and dried down, and digestion mix containing the restriction enzyme ApeKI (recognition site 5' G/CWGC 3', where W is A or T) is added to each sample. After the digest, a ligation master mix is added to each well. After the ligation is performed, samples are pooled and cleaned. A PCR step enriches ligated fragments and adds the sequences needed for binding to the flow cell.

In order to produce DNA fragments with a different adapter sequence on each end as is required for Illumina sequencing, the Elshire approach uses two different adapters, one that incorporates a 4–8 bp barcode and is specific for each individual, and another "common" adapter that is used by all individuals. Both adapters contain the ApeKI-specific overhang CWG. "Forward" and "reverse" oligos comprising each adapter are pre-annealed to one another by heating and then slowly cooling them (see Protocol). The lengths of the barcodes are varied to avoid all clusters reaching the cutsite on the same cycle near the beginning of the sequencing run, which can impede proper calibration of the sequencer.

In collaboration with the Rokhsar lab, I sought to use Illumina sequencing in concert with the strategies noted above to discover SNPs in the *X. tropicalis* genome, construct a genetic map, and apply that map to improve the genome assembly. An initial pilot using the Elshire approach suggested that it was not appropriate for *Xenopus*, so I designed a set of Y-shaped

adapters for a different restriction enzyme (Figures 2.1 and 2.2). In combination with these adapters, I achieved a reduced representation of the genome by sequencing a size fraction from a restriction digest, and I multiplexed up to 48 samples per flowcell lane using a different barcoded adapter for each individual (Figure 2.1).

Using this approach, I have successfully sequenced five multiplex libraries for *X. tropicalis* comprising the P0s, F1s, and F2s from a cross between the Nigerian and ICB strains, the two strains commonly used for mapping in this species (Khokha et al., 2009). We also adapted this method to *X. laevis*, successfully sequencing two multiplex libraries comprising F2 and backcross tadpoles from a cross between inbred strains.

All computational analysis of the data described in this chapter was performed by Therese Mitros.

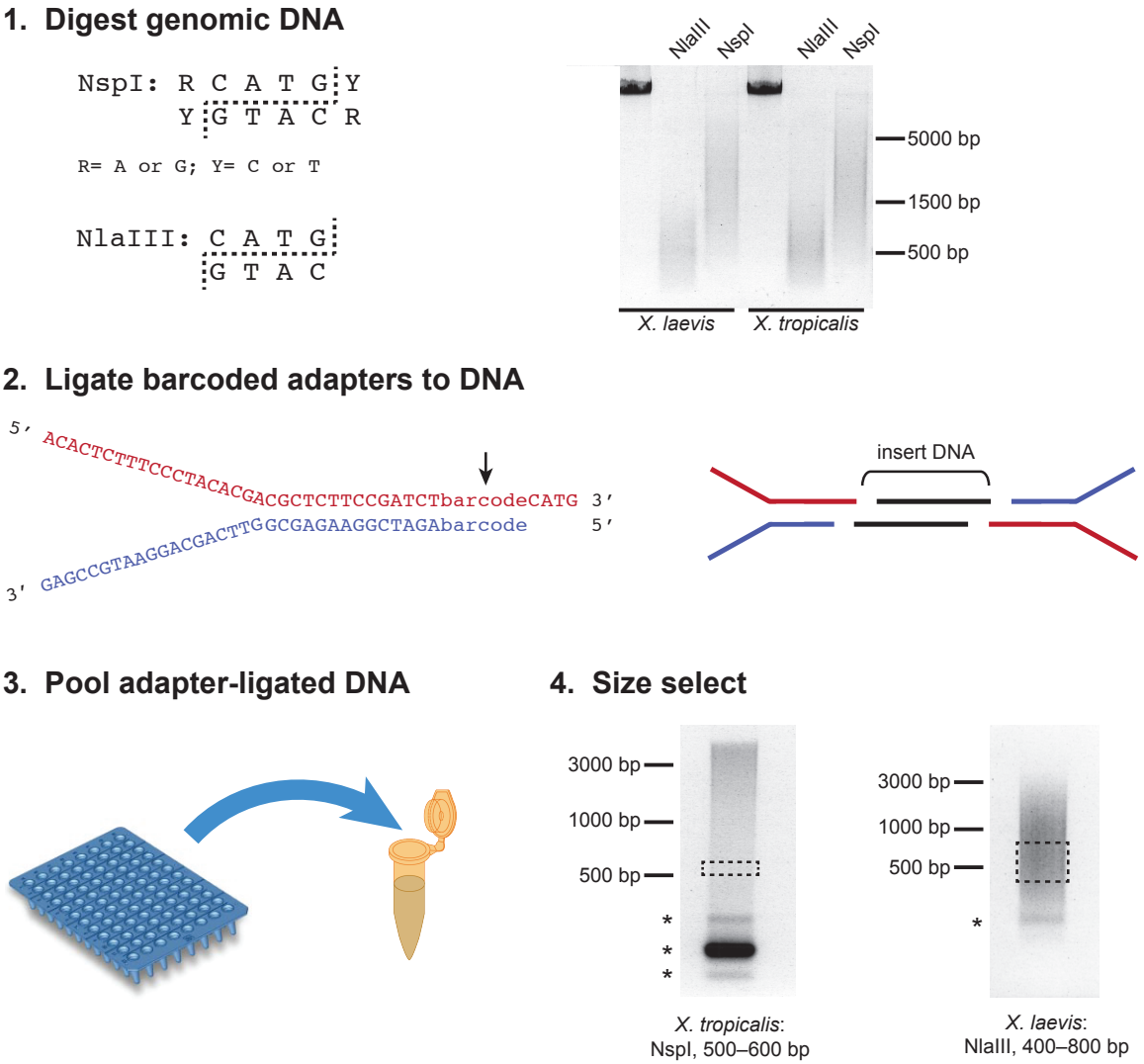


Figure 2.1. Experimental approach for GBS. (1) DNA is fragmented via restriction digest. NspI and NlaIII share the same overhang, but NlaIII is a more frequent cutter. (2) Adapters are ligated onto the ends of each DNA fragment. Y-shaped adapter sequences include sample-specific barcodes (arrow) and NspI/NlaIII-specific overhangs. (3) After adapter ligation, up to 48 samples are pooled. (4) Size selection is performed via agarose gel on adapter-ligated and pooled samples. Asterisks indicate adapter-mers. (5) After the PCR, each DNA fragment has the barcode on both ends, and a different adapter sequence on each end. The sequences represented in orange and green are added on during the PCR and facilitate binding to the flow cell.

1. Y-shaped adapters bind DNA via sticky-end ligation

2. PCR

PE1 5′CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT is complementary to blue sequence above
PE2 5′ AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT is identical to red sequence above, and isn't usable until the second cycle

3. Paired-end Sequencing

paired "read 1" seq primer: 5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT 3' matches red adapter sequence.
paired "read 2" seq primer: 5' CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT 3' is complementary to blue adapter sequence.

Figure 2.2. Adapter, PCR primer, and sequencing primer sequences. (1) When the Y-shaped adapter is ligated to both ends of the DNA, each end of each DNA strand has a different adapter sequence. (2) During the first PCR cycle, PCR primer PE1 primes a complementary copy of each strand from the blue end, adding length to that end, through to the end of the red sequence. During the second cycle, PE2 binds to the new strand from the red end, adding sequence on to that end. After the PCR, each double-stranded DNA fragment has a different adapter sequence on each end, and can bind the flow cell. (3) The sequencing primers are each complementary to one end of the amplified DNA fragment.

II. Genotyping by sequencing in *X. tropicalis* using the enzyme ApeKI

For our *X. tropicalis* genotyping by sequencing (GBS) libraries, we decided to use frogs from a mapcross for a recessive pigmentation mutant called *gray* (Figure 2.3) (Lane). Any mapcross between strains theoretically would have been usable for making a genetic map, but in this case we knew the color as well as the sex for each individual for the F2 DNAs isolated from adults. Thus we hoped to use the GBS data for mapping the *gray* and sex-determining loci as well, and to develop a protocol that could be used for mapping other mutations. We first tried a version of the protocol described in Elshire et al. Detailed parameters for all libraries I constructed are listed in Table 2.1.

For this ApeKI pilot experiment, we used only the P0s and F1s from our mapcross, with each individual's DNA distributed among a subset of the 48 wells (15 wells for the ICB P0, and 11 for each of the others). I followed a procedure almost exactly the same as the Elshire protocol (Table 2.1) (Elshire et al., 2011). Using 1 μL of pooled ligation as PCR template yielded a series of peaks rather than the desired single smooth peak; using 2 μL of template yielded a similar result (Figure 2.4). Performing a second round of PCR on the 1 μL library yielded a more even size distribution, and this library (Figure 2.4, ApeKI-1 reamp) was sequenced via 76 bp single-end reads, the longest reads available at the time. Longer sequence reads are desirable to optimize the alignment of the reads to the reference genome, and to increase the likelihood of finding a SNP in a given read. 47–64% of the reads mapped to the *X. tropicalis* genome assembly version 4 (Table 2.2). As one might expect given that the reference genome was sequenced from a frog of the Nigerian strain, the lowest percentage of reads from the ICB strain mapped to the reference.

Several attempts were made to construct an F2 library using 47 F2s. The most successful incorporated a size selection step of approximately 220–500 bp after pooling, as well as two rounds of PCR, but this nonetheless contained a series of peaks similar to those seen previously (Figure 2.4, ApeKI-S14), and was never sequenced. Based on the multiple peaks seen in the ApeKI libraries as well as the low mapping percentage of sequenced reads, we inferred that ApeKI may cut in a repetitive region of the *X. tropicalis* genome, and we decided to alter our approach to use a different restriction enzyme.

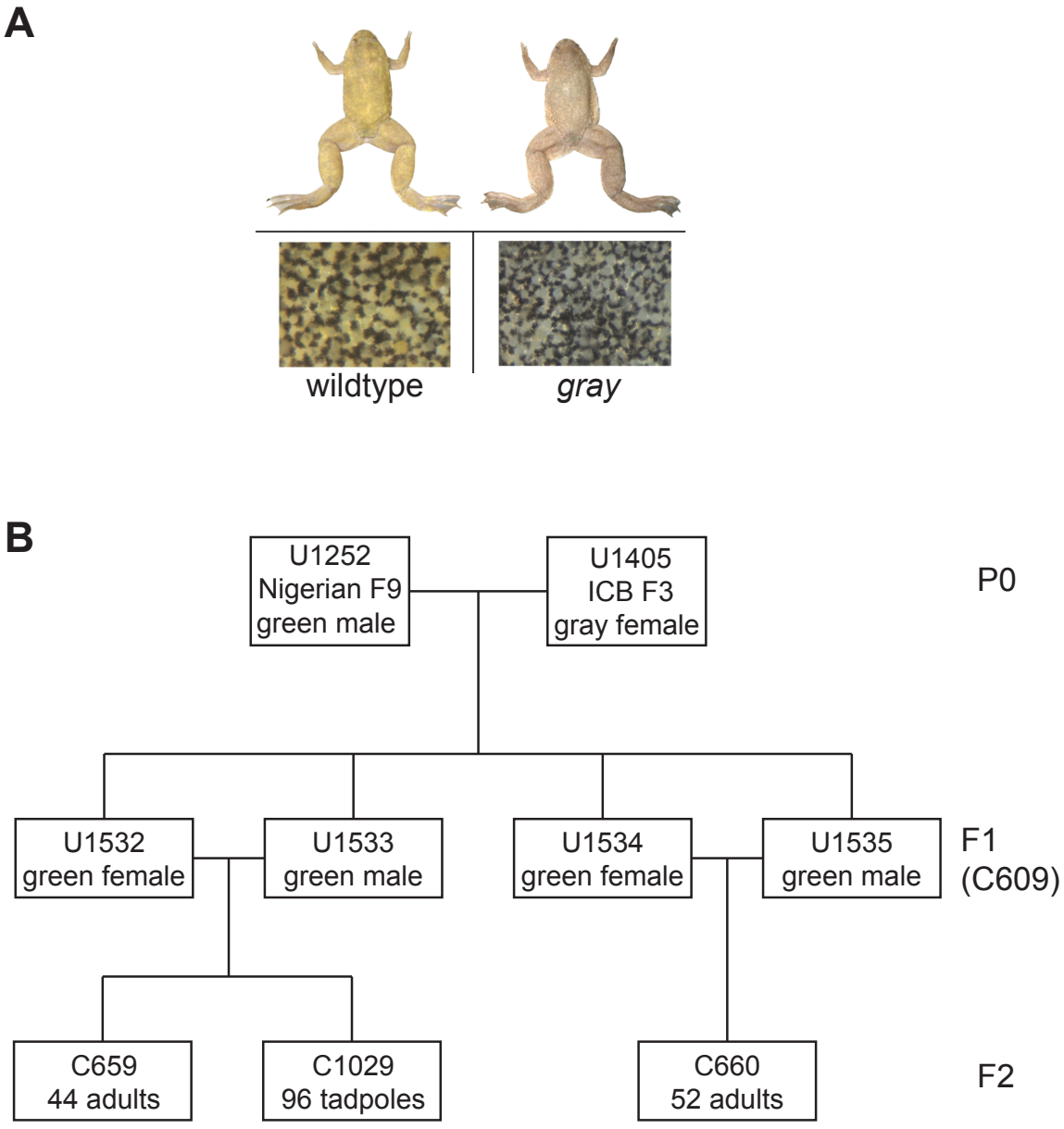


Figure 2.3. ***X. tropicalis*** **frogs genotyped by sequencing.** (A) Frogs homozygous for the *gray* mutation, a recessive Mendelian background mutation that arose in the ICB strain, lack yellow pigment (adapted from Lane et al., in preparation). (B) We sequenced frogs from a mapcross for the *gray* mutation. The numbers of F2 individuals sequenced in NspI libraries are indicated. For frogs whose DNA was obtained from an adult, we recorded phenotype (color and sex) information. A number preceded by U is an individual identifier for a frog; a number preceded by a C is a clutch number.

Table 2.1. Preparation parameters for Illumina libraries.

Library	Submitted as	Location in lab book	Samples	Digest Enzyme	DNA in digest or sheared	Digest or shearing	Final adapter conc in ligation	Ligation	Pooling	Size selection	PCR	Post-PCR cleaning	Sequencing
ApeKI-1, reamp	XEN-R	#3 p. 1–7	U1405, U1252, U1534, U1535; in 48 wells	ApeKI	30 ng/well + 10 µL adapter, dry down	10 µL master mix/well = 1.9–3.6 U/sample. 72°C 2 hours, cool to 4°C.	barcoded and common adapter, each 0.028 µM	In 50 µL: digest + 1.6 µL NEB T4 ligase & 5 µL NEB 10X T4 ligase buffer. 22°C 60 min, 65°C 30 min, cool to 4°C.	5 µL/well = 3 ng/sample = 144 ng total. Clean with Qiaquick PCR cleanup kit.	none	1 µL template, 14 cycles, then use 1 µL of that for a second round of PCR. In 50 µL 1X HF buffer, with 1 µL Phusion polymerase.	With Qiaquick PCR purification kit. Elute with 30 µL EB.	76 bp single-end on GAII
ApeKI-S14	N/A	#3 p. 32–34, 44, 51	47 F2s from C660	ApeKI	30 ng/well +10 µL adapter, dry down	3.6 U/sample in 10 µL. 72°C for 2 hours, cool to 12°C.	barcoded and common adapter, each 0.028 µM	In 50 µL: digest + 1.37–1.45 µL Roche T4 ligase & 5 µL NEB 10X T4 ligase buffer. 24°C 60 min, 65°C 30 min, cool to 12°C.	5 µL/well = 3 ng/sample = 141 ng total, clean with Qiaquick PCR cleanup kit.	~220–500 bp on 2% agarose, 73V. Gel extract with Qiaquick gel extraction kit, elute with 30 µL EB.	13 µL template, 14 cycles. In 50 µL 1X HF buffer, with 1 µL Phusion polymerase, 0.2 µM each primer, 0.2 mM each dNTP.	With Qiaquick PCR purification kit. Elute with 30 µL EB.	N/A
NspI-1	NspI	#3 p. 146–147, 150, 154–156	P0s; F1s U1534 & U1535; 8 F2s	NspI	P0s & F1s: 500 ng each. F2s: 3 µg each.	P0s & F1s (performed by Lisa Brunet): 500 ng DNA + 5U NspI in 25 µL. 2 hrs 37°C, then 65°C 20 min. F2s: 3 µg DNA + 27U NspI in 150 µL, 37°C overnight. ~147 µL purified with Minelute PCR purification kit, eluted with 25 µL EB.	0.5 µM	Performed by Lisa Brunet. In 50 µL: 0.5 ug digested DNA + adapter + 2 µL NEB T4 ligase & 5 µL ligase buff. 1 hour 22°C, 65°C 30 min. Then cleaned with Qiaquick or MinElute PCR purification kit to concentrate before gel.	N/A	Twice, performed by Lisa Brunet. P0s and F1s ~450–650bp; F2s 500–600 bp.	Performed by Lisa Brunet. In 100 µL, primers 1 µM final conc each, and 0.25 mM each dNTP. 15 cycles of Elshire et al. PCR program.	0.7X AMPure beads (of 85 µL remaining PCR), eluted with 30 µL H_2O.	100 bp single-end on HiSeq

Table 2.1, continued.

NspI-2	JBL 001	#4 p. 23–36	44 F2s from C660 blood preps	NspI	500 ng each from 1:5 dilutions	DNA + 10 U NspI in 50 µL. 4h 37°C, 20min 65°C, cool to 4°C. Froze O/N. Quantitate with Picogreen after.	0.25 µM	In 50 µL: 200 ng digested DNA + adapter, 5 µL NEB 10X T4 ligase buffer, 2 µL NEB T4 ligase. 2 hrs 22°C, 30 min 65°C, cool to 4°C.	5.5 µL/well = 22 ng/well = 968 ng total. Clean with MinElute kit.	Twice, performed by Lisa Brunet. 500–600 bp.	Performed by Lisa Brunet.	2 rounds with 0.7X AMPure XP beads, eluted with 30 µL H2O	100 bp paired-end on HiSeq
NspI-3	JBL 002	#4 p. 41-50	44 F2s from blood preps: 7 from C660 and 37 from C659. U1532 & U1533 prepared in parallel	NspI	500 ng each F2, all but one from 1:5 dilution; 1 µg each F1	F2s: 10U NspI in 50 µL. F1s: 20U NspI in 50 µL. 4 hrs or O/N 37°C, 20 min 65°C, cool to 4°C. Quantitate F2s with Picogreen after. Clean F1s with Minelute, elute with 25 µL H2O	F2s: 0.2 µM F1s: 0.54 µM	F2s: in 50 µL, 200 ng digested DNA + adapter, 5 µL NEB 10X T4 ligase buffer, 2 µL NEB T4 ligase. F1s: in 46 µL, all of cleaned digest + adapter, 3 µL NEB T4 ligase. 2 hrs 22°C, 30 min 65°C, froze.	F2s: 5.5 µL/well = 22 ng/sample = 968 ng total, clean with MinElute kit. F1s: clean only.	Twice. 2% agarose, ~140V, 500–600bp. Gel extract with MinElute kit, elute with 33 µL EB.	In 50 µL: 10 µL template, primers 0.02 µM each final, dNTPs 0.4 mM each final, 1 µL Phusion polymerase. 15 cycles.	With AMPure XP beads.	100 bp paired-end on HiSeq
NspI-4	JBL 003	#4 p. 103–112	48 F2 tadpoles from C1029	NspI	150 ng each, Picogreen quantitated	In 30 µL: 150 ng DNA, 3U NspI. 9 hrs 37°C, 20 min 65°C, cool to 4°C	adapter #1–4: 0.1 µM, #5–48: 0.125 µM	In 40 µL: 100 ng digested DNA, adapter, 4 µL NEB 10X T4 ligase buffer, 1 µL NEB T4 ligase. 2 hrs 22°C, 30 min 65°C, cool to 4°C.	8.4 µL/well (21 ng), except #328 & 332 (17 µL/well). ~1 µg total. Clean with MinElute kit, freeze.	Twice. 2% agarose, ~140V. 500–600 bp. Gel extract with MinElute kit, elute with 33 µL EB.	In 50 µL: 15 µL template, primers 0.02 µM each final, dNTPs 0.4 mM each final, 1 µL Phusion. 15 cycles.	0.7 volumes Ampure XP beads, elute with 20 µL H_20. Used Speedvac to concentrate remaining ~16.5 µL to <8 µL.	100 bp paired-end on HiSeq

Table 2.1, continued.

NspI-5	JBL 004	#4 p. 113–114, 118–124	48 F2 tadpoles from C1029	NspI	150 ng each, Picogreen quantitated	In 30 µL: 150 ng DNA, 3U NspI. 4 hrs 37°C, 20 min 65°C, cool to 4°C	adapter #1–4: 0.1 µM, #5–48: 0.125 µM	In 40 µL: 100 ng digested DNA, adapter, 4 uL NEB 10X T4 ligase buffer, 1 µL NEB T4 ligase. 16 hrs 16°C, 30 min 65°C, to 4°C.	8.4 µL/well (21 ng), except #357 & 369 (17 µL/well). ~1 µg total. Clean with MinElute kit.	Twice, 2% agarose, ~140V. 500–600bp. Gel extract with MinElute kit, elute with 32 µL EB.	In 50uL: 15 µL template, primers 0.2 µM each final, dNTPs 0.4 mM each final, 1 µL Phusion. 15 cycles.	0.7 volumes Ampure XP beads, elute with 30 µL H20.	100 bp paired-end on HiSeq
laevis F strain TruSeq	JBL 005	#4 p. 139–144, 148–149	X. laevis F strain female, DNA from blood, lysed in Robert lab. Dilution "1:5 C" cleaned, resuspended in H_20.	N/A	2200 ng, Picogreen quantitated, in 110 µL	Covaris shearing: 105 µL, used 500 bp protocol. Based on size selection gel, average size was more like 300 bp.	unknown	Used 50 µL sheared DNA and followed TruSeq protocol V2 for end repair, cleaning, A-tailing, ligation (adapter #11), cleaning.	N/A	Ethidium-stained 2% agarose, 120V. 700–800 bp. Gel extract with MinElute kit, elute with 32.5 µL RSB from kit	17uL size selected/gel extracted DNA, 3uL H20, 5uL PPC, 25uL PMM. Truseq PCR protocol, except 40 sec extension; 10 cycles.	per kit instructions, elute with 25uL RSB.	100 bp paired-end on HiSeq
X. laevis F/J males	JBL 006	#5 p. 1–9	"Frog 1A" and "Frog 1B": laevis F/J hybrid males from group A and B, respectively. DNA from blood	NlaIII	1 µg, Picogreen quantitated	in 50 µL: 1 µg DNA, 20U NlaIII. 8 hrs 37°C, 20 min 65°C. Then clean/concentrate with MinElute PCR purification kit, elute with 25 µL H_20.	0.32 µM	In 50 µL: 940 ng of digested DNA (cleaned with Minelute), adapter, 4 µL NEB 10X T4 ligase buffer, 3 µL NEB T4 ligase. 2 hrs 22°C, 30 min 65°C. Froze.	None, but clean with MinElute kit to concentrate.	Twice. 2% agarose, ~140V. 400–800 bp. Gel extract with MinElute kit, elute with 32 µL EB.	In 50 µL: 15 µL template, primers 0.2 µM each final, dNTPs 0.4 mM each final, 1 µL Phusion. 10 cycles.	0.7 volumes AMPure beads	150 bp paired-end on HiSeq, two lanes.
X. laevis F/JxJ (back cross)	JBL 007	#5 p. 1–9	39 F/JxJ tadpoles lysed in Robert lab	NlaIII	100 ng each, Picogreen quantitated	in 20 µL: 100 ng DNA, 2U NlaIII. 7 hours 37°C, 20 min 65°C	0.125 µM	In 40 µL: digest + adapter, 4 µL NEB 10X T4 ligase buffer, 1 µL NEB T4 ligase. 2 hrs 22°C, 30 min 65°C. Froze.	10.26 µL/well (25.64 ng), ~1 µg total. Clean with MinElute kit.	Twice, 2% agarose, 140V. 400–800 bp, extract with MinElute, elute with 32 µL EB.	In 50 µL: 15 µL template, primers 0.2 µM each final, dNTPs 0.4 mM each final, 1 µL Phusion. 10 cycles.	0.7 volumes AMPure beads	150 bp paired-end on HiSeq

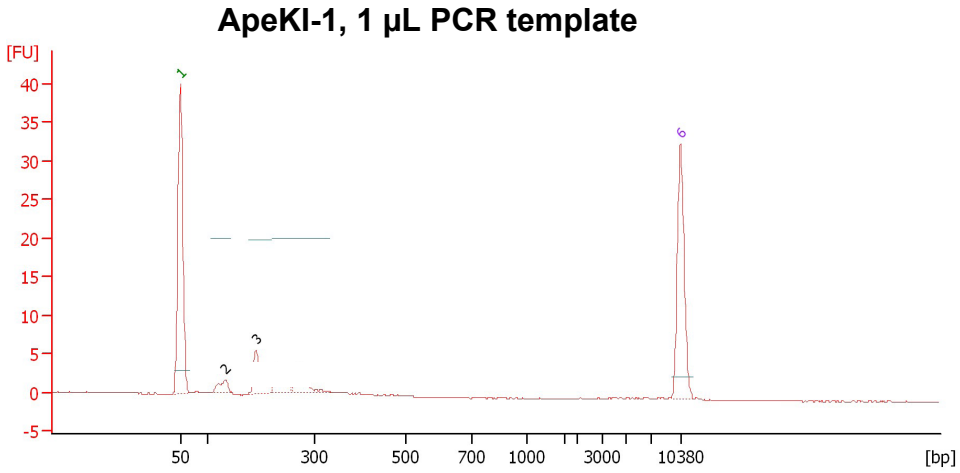
Table 2.1, continued.

X. laevis F/JxF/J (F2s)	JBL 008	#5 p. 1–9	29 F/JxF/J tadpoles lysed in Robert lab	NlaIII	18–50 ng each, Picogreen quantitated. Varying amounts of degradation.	In 20 µL: 18–50 ng DNA, 1U NlaIII. 7 hours 37°C, 20 min 65°C	0.0625 µM	In 40 µL: digest + adapter, 4 µL NEB 10X T4 ligase buffer, 0.6 µL NEB T4 ligase. 2 hrs 22°C, 30 min 65°C. Froze.	18 ng/well (varying volumes), total 522 ng. Clean with MinElute kit.	Twice, 2% agarose, 140V. 400–800 bp, extract with MinElute, elute with 32 µL EB. Couldn't see DNA on gel.	in 50 µL: 15 µL template, primers 0.2 µM each final, dNTPs 0.4 mM each final, 1 µL Phusion. 15 cycles.	0.7 volumes AMPure beads	150 bp paired-end on HiSeq
***X. tropicalis* Nigerian F13 males: TruSeq**	JBL 014	#5 p. 24–32	Toes from 12 Nigerian F13 males, pooled	N/A	183.33 ng/frog, Picogreen quantitated: 2200 ng total in 110 µL	Covaris shearing: 105 µL, used 500 bp protocol. Based on size selection gel, average size was approx 400–500 bp.	unknown	Used 50 µL sheared DNA and followed TruSeq protocol V2 for end repair, cleaning, A-tailing, ligation (adapter #5), cleaning.	N/A	EtBr-stained 2% agarose, 140V. 500–600 bp. Gel purify with Minelute, elute with 25 µL EB	20uL size selected/gel extracted DNA, 5uL PPC, 25uL PMM. Truseq PCR protocol, except 40 sec extension; 10 cycles.	Per kit instructions. Elute with 32.5 µL RSB.	100 bp paired-end on HiSeq
***X. tropicalis* Nigerian F13 females: TruSeq**	JBL 015	#5 p. 24–32	Toes from 8 Nigerian F13 females, pooled	N/A	275 ng/frog, Picogreen quantitated: 2200 ng total in 110 µL	Covaris shearing: 105 µL, used 500 bp protocol. Based on size selection gel, average size was approx 400–500 bp.	unknown	Used 50 µL sheared DNA and followed TruSeq protocol V2 for end repair, cleaning, A-tailing, ligation (adapter #6), cleaning.	N/A	EtBr-stained 2% agarose, 140V. 500–600bp. Gel purify with Minelute, elute with 25 µL EB.	20 µL size selected/gel extracted DNA, 5 µL PPC, 25 µL PMM. Truseq PCR protocol, except 40 sec extension; 10 cycles.	Per kit instructions. Elute with 32.5 µL RSB.	100 bp paired-end on HiSeq
NlaIII-B: F/JxJ and F/JxF/J tadpoles	JBL 016	#5 p. 34–41	11 F/JxJ and 30 F/JxF/J tadpoles lysed in Robert lab	NlaIII	100 ng each, from Picogreen quantitated 30 ng/µL dilutions	In 20 µL:100 ng DNA, 2U NlaIII. 4 hours 37°C, 20 min 65°C	0.125 µM	In 40 µL: digest + adapter, 4 µL NEB 10X T4 ligase buffer, 1 µL NEB T4 ligase. 2 hrs 22°C, 30 min 65°C. Froze.	9.76 µL (24.4 ng)/well, total 1 µg. Clean with MinElute.	Twice. 2% agarose, 140V. 400–800 bp. Gel extract with MinElute, elute with 22 µL EB.	In 50 µL: 10 µL template, primers 0.2 µM each final, dNTPs 0.4 mM each final, 1 µL Phusion. 10 cycles with extension time 45 sec.	0.7 volumes AMPure beads	100 bp paired-end on HiSeq

Table 2.1, continued.

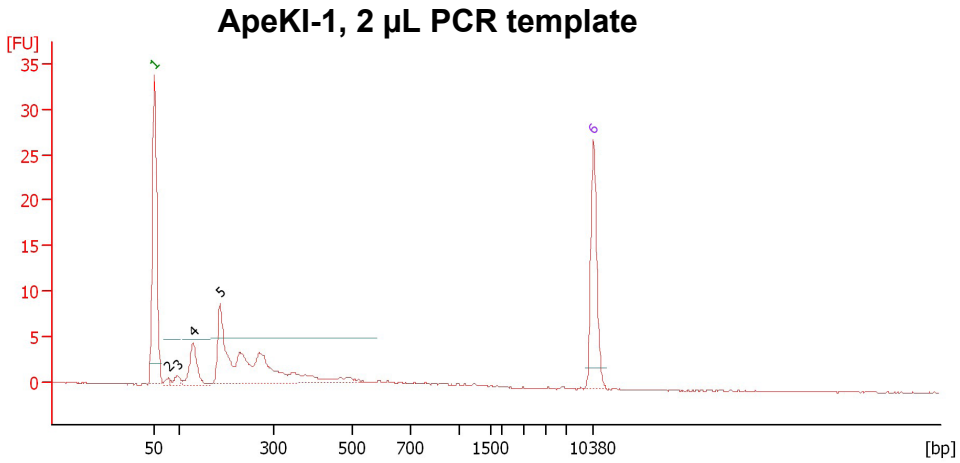
ICB F7 males: TruSeq	JBL 017A	#5 p. 42, 44–49	Toes from 6 F7 ICB males, sent from Khokha lab. Pooled.	N/A	366.67 ng/frog, from Picogreen quantitated 1:3 dilutions	Covaris shearing: 105 µL, used 500 bp protocol. Based on size selection gel, average size was approx 400–500 bp.	unknown	Used 50 µL sheared DNA and followed TruSeq protocol V2 for end repair, cleaning, A-tailing, ligation (adapter #2), cleaning.	N/A	EtBr-stained 2% agarose, 140V. 500–600bp. Gel purify with Minelute, elute with 25 µL EB	20 µL size selected/gel extracted DNA, 5 µL PPC, 25 µL PMM. Truseq PCR protocol, except 40 sec extension; 10 cycles.	Per kit instructions. Elute with 32.5 µL RSB.	150 bp paired-end on HiSeq
ICB F7 females: TruSeq	JBL 017B	#5 p. 42, 44–49	Toes from 6 F7 ICB females, sent from Khokha lab. Pooled.	N/A	366.67 ng/frog, from Picogreen quantitated 1:3 dilutions	Covaris shearing: 105 µL, used 500 bp protocol. Based on size selection gel, average size was approx 400–500 bp.	unknown	Used 50 µL sheared DNA and followed TruSeq protocol V2 for end repair, cleaning, A-tailing, ligation (adapter #4), cleaning.	N/A	EtBr-stained 2% agarose, 140V. 500–600bp. Gel purify with Minelute, elute with 25 µL EB.	20 µL size selected/gel extracted DNA, 5 µL PPC, 25 µL PMM. Truseq PCR protocol, except 40 sec extension; 10 cycles.	Per kit instructions. Elute with 32.5 µL RSB.	

Figure 2.4. Bioanalyzer results for ApeKI libraries.



Peak table :

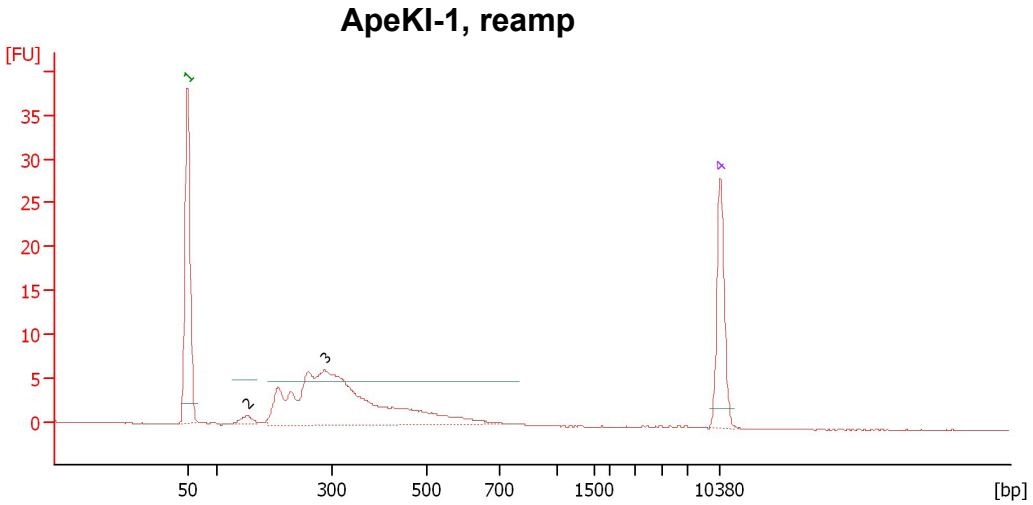
Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	50	8.30	251.5	Lower Marker
2		132	0.47	5.3	
3		189	1.26	10.1	
4		232	0.56	3.7	
5		273	0.50	2.8	
6	▶	10,380	4.20	0.6	Upper Marker



Peak table :

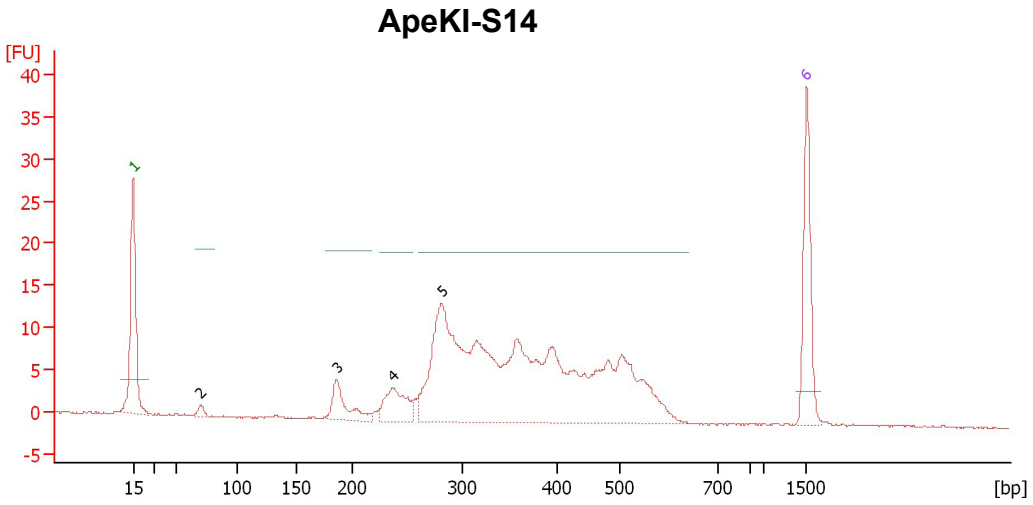
Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	50	8.30	251.5	Lower Marker
2		78	0.14	2.7	
3		97	0.33	5.2	
4		131	1.57	18.2	
5		188	7.89	63.7	
6	▶	10,380	4.20	0.6	Upper Marker

Figure 2.4, continued.



Peak table :

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	50	8.30	251.5	Lower Marker
2	153	0.38	3.8	
3	287	13.50	71.3	
4	10,380	4.20	0.6	Upper Marker



Peak table :

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	70	0.19	4.1	
3	186	0.99	8.1	
4	238	1.13	7.2	
5	281	18.50	99.8	
6	1,500	2.10	2.1	Upper Marker

Table 2.2. Reads mapped to the *X. tropicalis* reference from ApeKI-1 and NspI-1.

Frog	Generation	# Reads: ApeKI-1	% mapped: ApeKI-1
U1405	P0 ICB	3,350,013	47.4%
U1252	P0 Nigerian	5,227,674	57.3%
U1534	F1	2,472,622	56.6%
U1535	F1	2,666,280	64.0%

Frog	Generation	# Reads: NspI-1	% mapped: NspI-1
U1405	P0 ICB	3,222,776	87.6%
U1252	P0 Nigerian	2,704,358	93.3%
U1534	F1	2,687,274	91.6%
U1535	F1	4,079,842	90.8%
#34	F2	1,449,580	90.9%
#35	F2	1,720,493	90.5%
#149	F2	1,267,787	89.6%
#150	F2	1,126,628	90.0%
#31	F2	2,118,366	90.0%
#32	F2	1,541,846	90.3%
#5	F2	2,278,706	91.3%
#6	F2	2,266,053	91.1%

III. Genotyping by sequencing in *X. tropicalis* using the enzyme NspI

Early versions of this protocol were developed in consultation with Mustafa Khokha and Lisa Brunet.

IIIa. Protocol design and optimization

I designed a set of adapters for the restriction enzyme NspI, which has a six-base-pair recognition site with degeneracy outside the overhang [Figure 2.1(1)], so its frequency of cutting should be intermediate between that of an enzyme with a six-base-pair recognition site and that of an enzyme with a four-base-pair recognition site. *in silico* digests (performed by U. Hellsten and T. Mitros) of the *X. tropicalis* genome did not suggest that NspI would digest the *X. tropicalis* genome in a biased manner. It is worth noting, however, that the genome assembly available at the time, version 4, probably did not contain repetitive regions. Based on the *in silico* digest, we decided to size select for 500–600 bp fragments, which should represent approximately 2–3% of the *X. tropicalis* genome.

As described above, the Elshire approach uses two different adapters per sample, but since the two ends of each DNA fragment have the same overhang, that approach would result in only half of the fragments having a different adapter on each end and the rest would have one or the other adapter on both ends. The use of Y-shaped adapters avoids this problem, so I designed the NspI adapters based on the Illumina Y-shaped paired-end adapters (Figures 2.1 and 2.2)(Bentley et al., 2008). Non-reduced representation, or shotgun, Illumina sequencing uses shearing to fragment the DNA, and Klenow enzymes are used to fill in and then A-tail the ends of the fragments. The overhanging A on each end is then used for ligation to a Y-shaped adapter with an overhanging T. I included this T in my adapters as a TA pair because the sequencing primer is designed to bind adapters up to and including that base. The TA pair is followed by the barcode and the CATG overhang required for ligation to NspI-digested DNA fragments (Figure 2.2). Thus, depending on the length of the barcode, the first 4–8 bases of every sequence should be the barcode.

One reason that Elshire and colleagues may have designed a two-adapter system is that even in paired-end reads the barcode is only strictly required on one end, because the pairing of the reads allows for both ends to be associated with the barcode. If the barcode is only on one end of a given DNA fragment then a maximum of 8 bases is "lost" to barcode sequencing. This concern, however, is less relevant with the advent of longer reads. Furthermore, screening all reads for proper barcoding builds in an extra level of quality control.

I ordered an initial set of 12 NspI adapters (#1–12) using barcodes chosen from among those reported in Elshire et al. These were used for troubleshooting as well as in the first submitted library (barcode sequences are listed in Table 2.3 and Table 2.6). As per the Elshire protocol, the first set of oligos I ordered were not 5' phosphorylated. Although theoretically a phosphate group would be required on the 5' end of the "reverse" adapter oligo [in blue in Figure 2.1(2)], the ligation and subsequent amplification still somehow works without it (e.g. Elshire et al. 2011 and my ApeKI library XEN-R). I later ordered phosphorylated versions

of some of our adapter reverse oligos for testing. In my tests, I found that a minimum of 0.8 μM concentration of non-phosphorylated NspI adapter was required in a ligation containing 50 ng DNA, as assessed by effective and repeatable amplification using the adapter-specific PCR primers (data not shown). In contrast, a concentration of as little as 0.2 μM phosphorylated adapter was sufficient for effective ligation and amplification, and concentrations greater than 0.2 μM appeared inhibitory (Figure 2.5A). One might hypothesize that the inhibition might be derived from adapter-mers amplifying at the expense of adapters ligated to insert DNA, but curiously there was no evidence of amplified adapter-mer bands.

In the end, in consultation with Lisa Brunet, I decided to use phosphorylated reverse oligos in all adapters, and all of the NspI ligations used adapter concentrations similar to 0.1 μM per 100 ng of DNA (Table 2.1). The full set of adapters used 48 distinct barcodes chosen from among those reported in Elshire et al.; the barcode TAGCATGC, which contains the NspI recognition site, was modified to TAGCAGCC (Table 2.6). Adapter oligos were ordered from Eurofins MWG Operon with HPSF purification. Lisa and I wanted to ensure that the adapters were in excess in order to avoid DNA dimerization, which could confound mapping to the reference genome. I later confirmed that the adapters were in excess because adapter-mers were evident when the ligations were run on a gel for size selection [e.g. as in Figure 2.1(4)].

I decided to dispense with the initial drying down of the DNA and adapters described in the Elshire protocol because the concentrations of reagents made it unnecessary. Instead, Lisa and I simply added the same number of nanograms of DNA for each sample to each digest, and normalized the volumes using the appropriate amounts of water. We also decided to incorporate a heat-inactivation step into the digest (ApeKI can't be heat-inactivated), and to add the adapters just before the ligation step, in contrast to the Elshire protocol.

F2 DNAs prepared from adults were isolated from blood (Additional materials and methods), and tended to be viscous. An even distribution of reads between pooled individuals hinges on each individual contributing the same amount of DNA to the pool. Due to concerns that viscosity of DNAs from blood might impede precise quantitation, we inserted a second DNA quantitation step post-digest for pooled multiplex libraries made from blood DNAs (NspI-2 and NspI-3). For these libraries, approximately 500 ng of DNA was digested. In the post-digest quantitation of NspI-2, for example, the DNAs should have been approximately 10 ng/μL, but they ranged from 7 to 39 ng/μL. After re-quantitation, 200 ng were used in the ligation, and ligation volumes were normalized with water.

The Elshire protocol did not call for a size selection step, but I planned to incorporate a size selection in order to achieve the stringent reduced representation discussed above. During initial optimization trials, I tested various parameters by making libraries without any size selection, evaluating their relative merits by comparing relative amplification efficiencies (e.g. Figure 2.5A). Although amplification was evident, I was initally troubled by the lack of DNA in my desired size range of 500–600 bp (Figure 2.5B). However, the incorporation of a size selection step prior to the PCR amplification facilitated the specific amplification of DNA fragments in the desired size range (Figure 5B').

Lisa and I decided to perform two rounds of size selection to ensure the removal of any adapter-mers that might result from having included the adapters in excess. Repeating the size selection also may have improved its stringency by mitigating any gel-to-gel variability in fragment migration. We used the Qiagen MinElute kit for gel extraction, melting the gel at room temperature rather than 50°C to avoid GC bias (Quail et al., 2008).

Finally, I switched to using AMPure XP SPRI (solid-phase reversible immobilization; i.e. magnetic) beads (Beckman Coulter, Inc.) for the post-PCR cleaning because they are effective in removing any residual adapter-mers: 0.7 volumes removes DNA fragments less than 300 bp in length (Quail et al., 2008; Lennon et al., 2010).

Lisa Brunet performed the PCR reactions for the versions of NspI-1 and NspI-2 that were submitted for sequencing, because the amplification was more effective in her hands than mine, and we could not initially discern why (compare e.g. the Bioanalyzer results from NspI-2, PCR performed by Lisa; with NspI-3, PCR performed by me; Figure 2.6). There were two main differences between my PCR parameters, which were based on the Elshire protocol, and hers: she used a 100 μL volume and much more PCR primer (1 μM vs 0.02 μM final concentration; Table 2.1). When I used exactly her parameters, I puzzlingly got only what appeared to be a bright adapter-mer band, whereas with my parameters I got the expected band between 500 and 600 bp, although it was not as robust as it perhaps could have been (Figure 2.5C; NspI-3). I now suspect that the PCR machines I was using aren't configured to thermocycle 100 μL reaction volumes. Nonetheless, libraries NspI-3 and NspI-4, amplified with my PCR parameters, sequenced successfully, and one might argue that less PCR amplification is desirable given the potential for introduced bias. Finally, during the preparation of NspI-5, I tested the effect of increasing the amount of primer in my PCR reaction (by tenfold; Figure 2.5), and the amplification worked much better. Elshire et al. had titrated down the amount of primer and adapter they used because they wanted to avoid adapter-/primer-mer contamination in their non-size selected libraries. Given the effectiveness of AMPure beads in removing excess adapter and primer, however (Quail et al., 2008), there is no reason to allow either of these reagents to be limiting. Repeatable robust amplification allows for fewer PCR cycles to be used, reducing potential bias. I therefore reduced the number of PCR cycles to 10 for the *X. laevis* libraries.

IIIb. NspI library preparation and sequencing

For our pilot NspI library (NspI-1), we sequenced the P0s, the F1 parents of clutch #660 (C660), and 8 F2s from C660. Since it only comprised 12 individuals, we prepared each as a separately in parallel rather than pooling them. I performed the digests of the F2 individuals, and Lisa Brunet prepared the digests of the P0s and F1s as well as the ligations, size selections, and PCR amplifications. As described above, the F2s were size selected for 500–600 bp. We decided to select a broader range of approximately 450–650 bp for the P0s and F1s in order to ensure that data from these individuals would cover all loci sequenced from F2s. I purified the PCR products, and they were sequenced as 100 bp single-end reads.

The Bioanalyzer results for NspI-1 are shown in Figure 2.6. Based on QPCR quantitation, the 12 libraries were combined onto one lane by the sequencing facility as the following approximate percentages of total clusters: P0s and F1s 12.5% each, F2s at 6.25% each.

The read counts from NspI-1 for each individual are listed in Table 2.2. As expected, the ICB P0 individual had the lowest percentage (88%) of reads mapping to the reference, and the Nigerian P0 individual had the highest percentage (93%). The F1s and F2s yielded mapping percentages intermediate between those two values. Importantly, the mapping percentages from this library were markedly higher than those from the ApeKI pilot, and so we decided to proceed with libraries made using NspI.

The next two NspI libraries (NspI-2 and -3), comprising F2 DNAs from adult blood, were prepared as described above and in Table 2.1. Each comprised 44 F2 individuals from C660 or C659. We avoided assigning barcodes 1–4 to these F2s to allow for the mixing in of P0 or F1 libraries (which used barcodes 1–4, Table 2.3). The P0s were combined in with NspI-2 by the sequencing facility at 10% each of the total clusters, and the two previously sequenced F1s at 2% each. Individual libraries for the F1 parents of C659 (U1532 and U1533) were prepared in parallel with NspI-3, and were combined in with that library at approximately 8% each of the total clusters. Bioanalyzer results for these libraries are shown in Figure 2.6.

NspI-4 and NspI-5 together comprised 96 DNAs from lysed tadpoles from a recent mating between the F1 parents of C659 (Figure 2.3). Although these DNA samples were not associated with any phenotypic information, they represent a choice to sequence more individuals rather than obtain more coverage from previously sequenced individuals: low coverage can be tolerated from F2 individuals because neighboring markers can be binned and used to impute the genotype at a given marker (Davey et al., 2011). As tadpole DNAs are much less viscous than blood DNAs, the post-digest quantitation step was not performed. Bioanalyzer results for NspI-4 and NspI-5 are shown in Figure 6, and the comparison between them exemplifies the dramatic increase in amplification resulting from the optimization of my PCR parameters described above. Table 2.4 summarizes the F2 read counts from all five NspI libraries, and Table 2.6 lists the phenotype, barcode, and read count for each F2 individual.

Table 2.5 shows the total read counts for each of the NspI libraries. With the advent of Illumina's V3 chemistry, there was a marked increase in the read counts obtained from NspI-4 and NspI-5 over previous libraries. "Attributed reads" refers to the number of reads or read pairs attributed to an individual (paired-end reads are mapped as pairs, with a specific size range allowed between where the two reads align). For single-end sequenced libraries (i.e. NspI-1), this requires that the read contain the barcode as well as the cutsite (for NspI libraries, CATGY). For paired-end sequenced libraries, it was required that both reads in a pair contain the barcode and cutsite. This more stringent requirement may explain the reduced percentages of attributed reads in paired-end libraries relative to NspI-1. NspI-4 had a particularly high number of read pairs where the barcodes on the two reads did not agree, which may explain its particularly low rate of attribution overall and perhaps its high coefficient of variation among F2s.

Table 2.3 gives the read counts for the P0 and F1 individuals. A high amount of coverage from the P0s was desired in order to confidently define the Nigerian and ICB genotypes. Therese identified an initial list of 46,236 SNPs, where the Nigerian and ICB individuals were each homozygous and differed from the other. The F1s were used as a check for these SNPs, based on the assumption that if a given SNP is homozygous and different between Nigerian and ICB P0s, then the F1s should be heterozygous for that SNP. 798 SNPs were excluded based on apparent homozygosity in the F1s, leaving a list of 45,438 SNP markers.

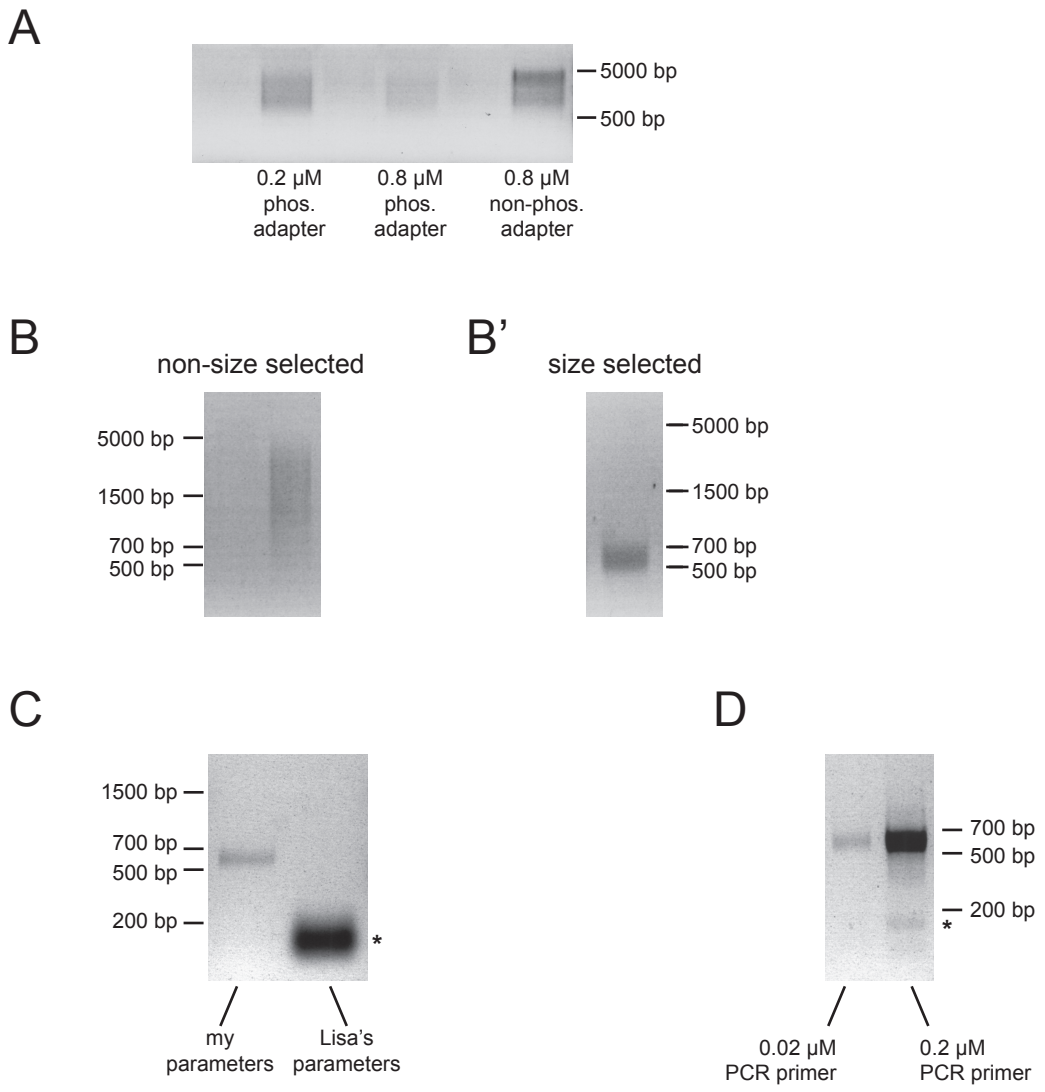


Figure 2.5. Troubleshooting NspI library construction parameters. (A) A greater concentration of non-phosphorylated than phosphorylated adapter is required for successful amplification. Although non-phosphorylated adapter concentrations of less than 0.8 µM were unreliable (data not shown), concentrations greater than 0.2 µM of phosphorylated adapter appear inhibitory. An equal amount of digested DNA is shown to the left of each amplification test. (B & B') Although not much DNA of the desired size amplified in non-size selected libraries, size selection allowed for the specific amplification of DNA fragments in a desired size range. The left lane in (B) is a digested-only DNA control. The amplified libraries in B and B' were made from the same digested DNA. (C) Although amplification using my PCR parameters was not initially robust, I was completely unsuccessful with Lisa's parameters. (D) Increased PCR primer concentrations allow for increased amplification of Illumina libraries. Asterisks mark probable adapter-mer bands (these PCR products had not been cleaned with AMPure beads).

Figure 2.6. Bioanalyzer results for *X. tropicalis* NspI libraries.

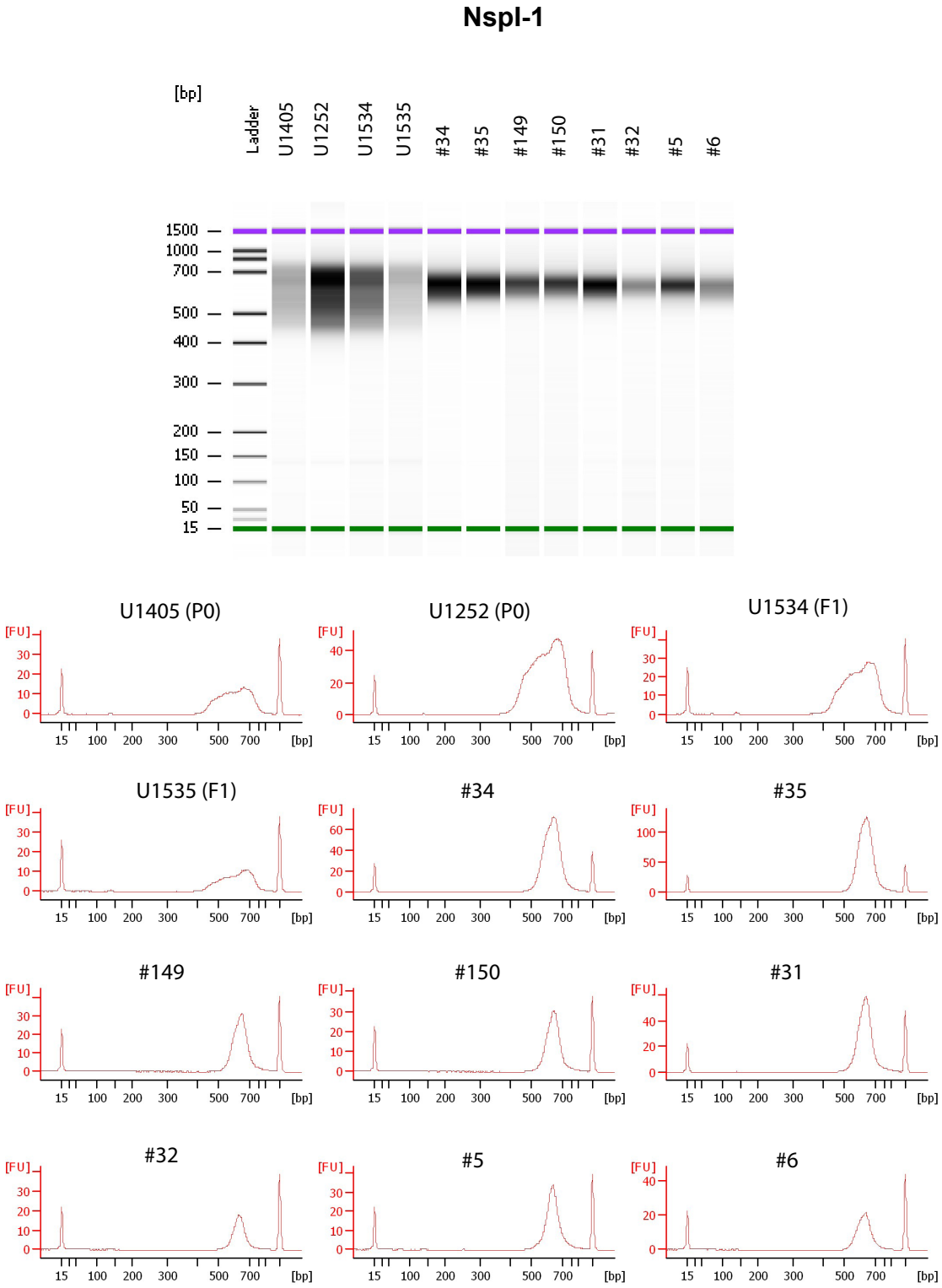
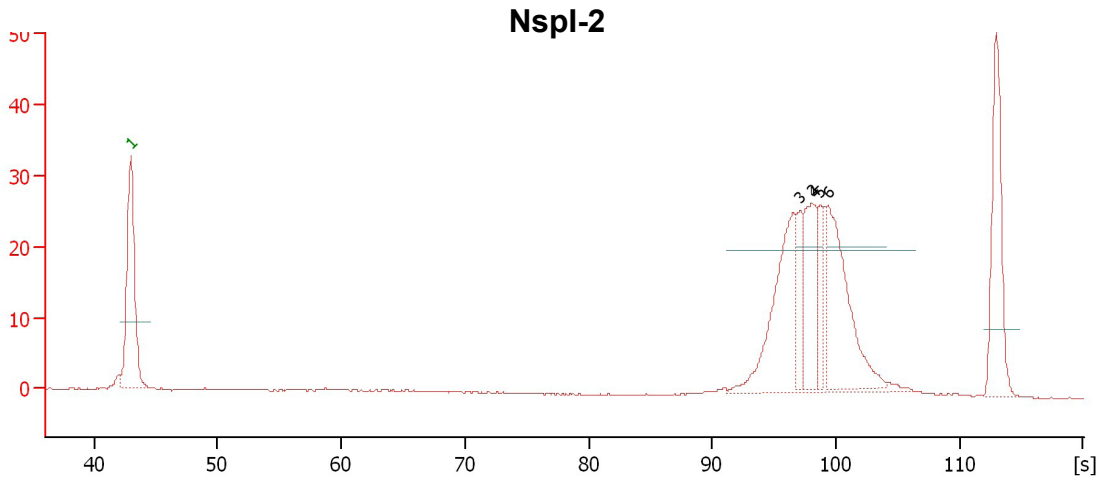
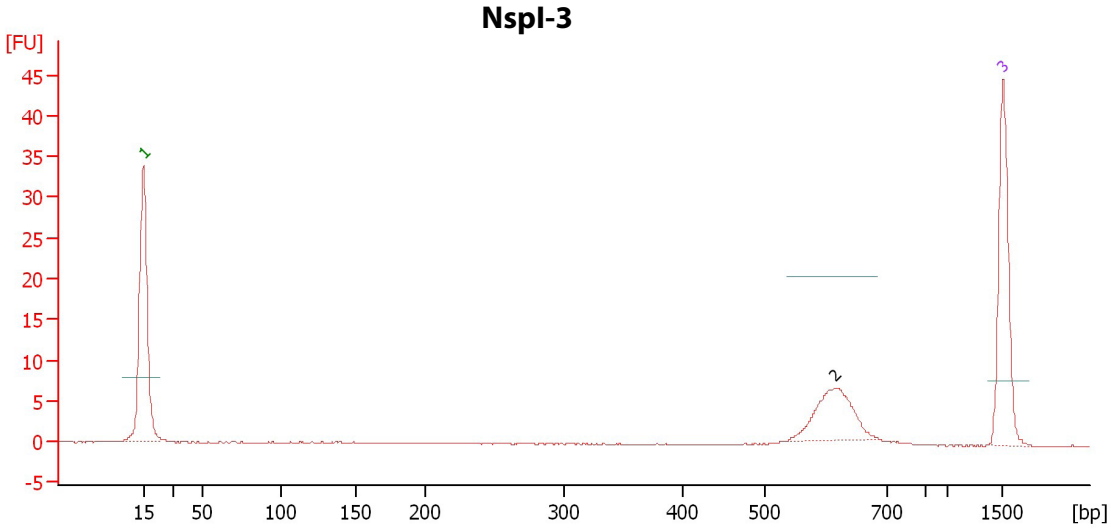


Figure 2.6, continued.



Peak table:

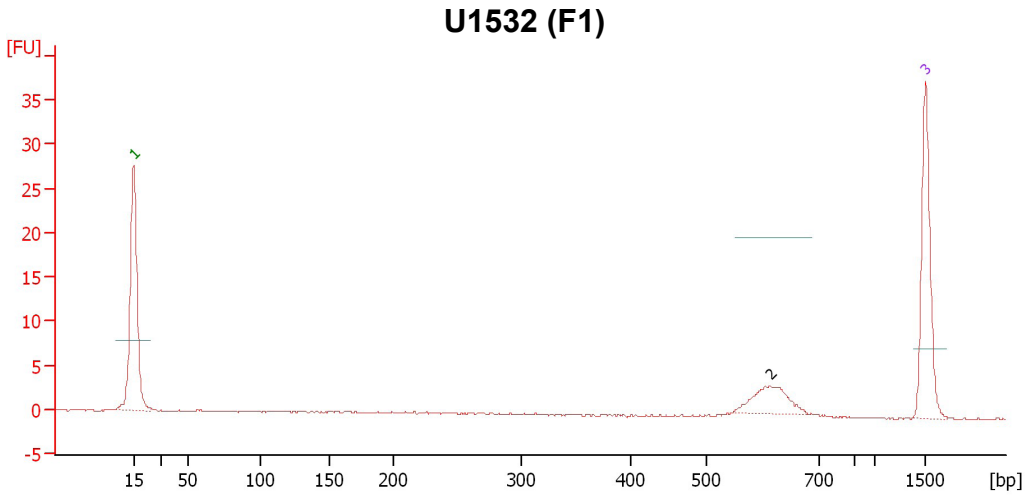
Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		596	9.43	23.9	
3		576	0.73	1.9	
4		600	1.58	4.0	
5		610	0.75	1.9	
6		623	2.89	7.0	
7	▶	1,500	2.10	2.1	Upper Marker



Peak table :

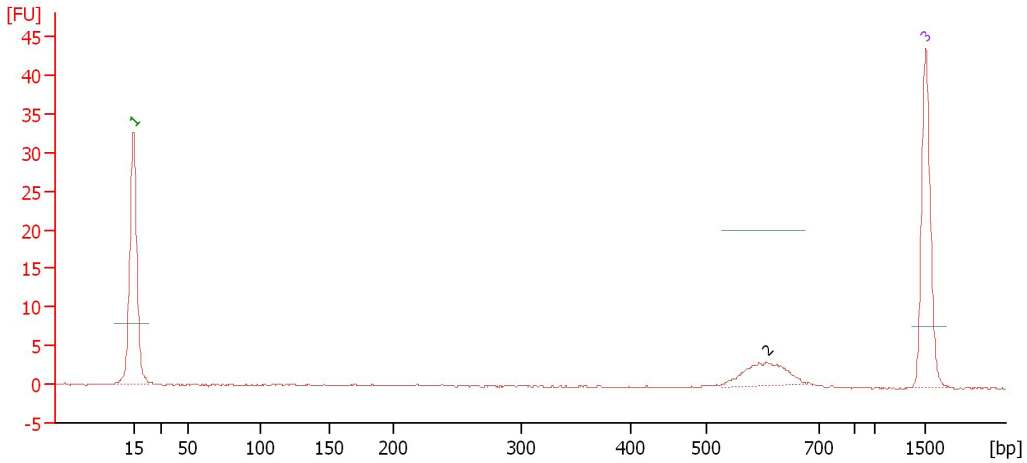
Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		615	1.31	3.2	
3	▶	1,500	2.10	2.1	Upper Marker

Figure 2.6, continued.



Peak table :

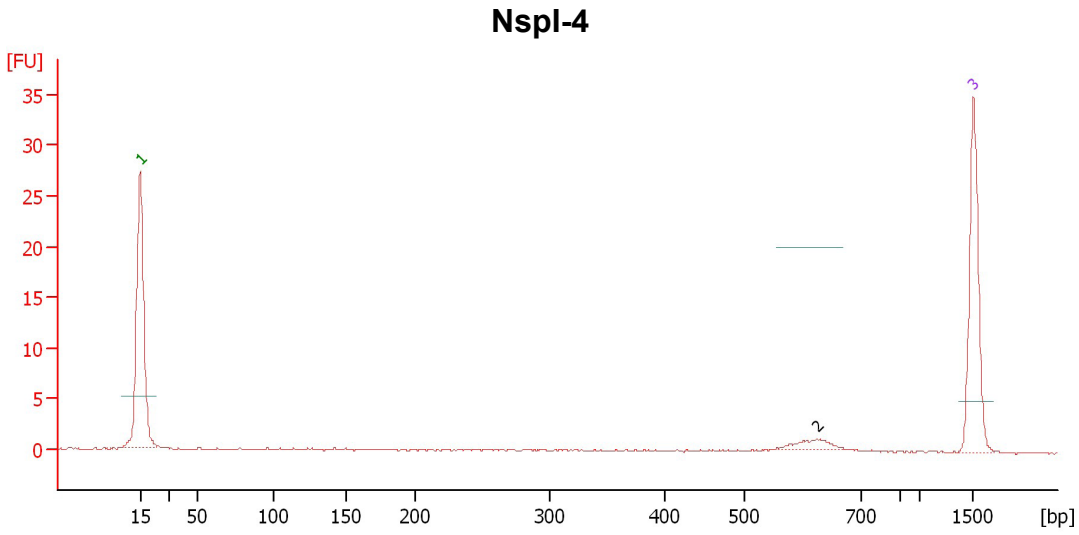
Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		613	0.85	2.1	
3	▶	1,500	2.10	2.1	Upper Marker



Peak table:

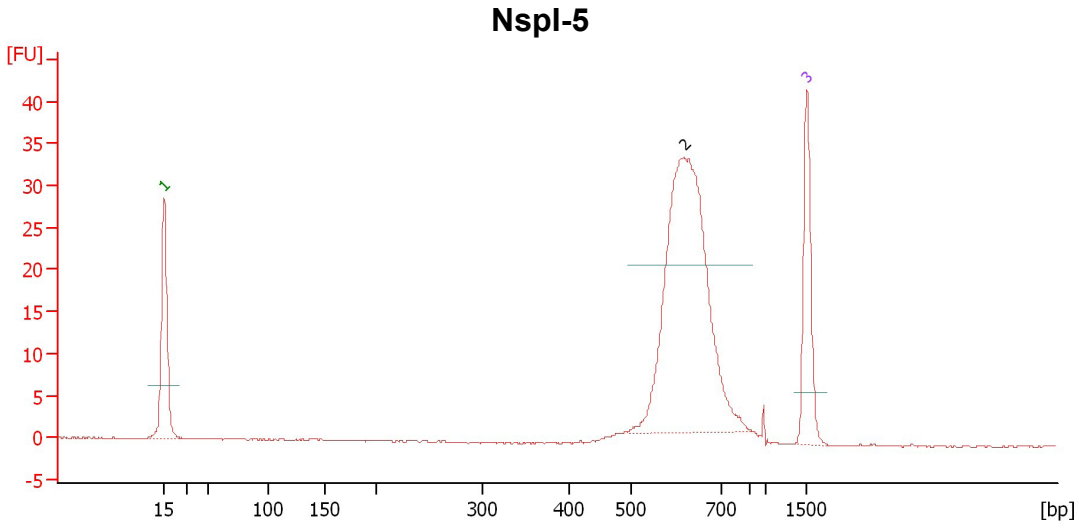
Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		607	0.77	1.9	
3	▶	1,500	2.10	2.1	Upper Marker

Figure 2.6, continued.



Peak table :

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		624	0.27	0.7	
3	▶	1,500	2.10	2.1	Upper Marker



Peak table :

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		618	10.73	26.3	
3	▶	1,500	2.10	2.1	Upper Marker

Table 2.3. Read counts for P0 and F1 individuals from *X. tropicalis* NspI libraries.

Frog	Generation	Color	Sex	Barcode number	Barcode sequence	NspI-1 reads	NspI-2 read pairs	NspI-3 read pairs	Total reads
U1405	P0 ICB	gray	female	1	acta	3,222,776	5,835,423		14,893,622
U1252	P0 Nigerian	green	male	2	tcacc	2,704,358	4,822,485		12,349,328
U1534	F1	green	female	3	ttctc	2,687,274	1,149,552		4,986,378
U1535	F1	green	male	4	gtatt	4,079,842	1,575,352		7,230,546
U1532	F1	green	female	3	ttctc			2,796,539	5,593,078
U1533	F1	green	male	4	gtatt			1,490,277	2,980,554

Table 2.4. Summary of F2 read counts from *X. tropicalis* NspI libraries.

Library	# F2s	Total reads attributed to F2 individuals	Average # reads per F2 individual	Range	Standard deviation	Coefficient of variation
NspI-1	8	13,769,459	1,721,182	1,126,628–2,278,706	452,053	0.26
NspI-2	44	86,968,994	1,976,568	1,247,090–2,847,744	320,234	0.16
NspI-3	44	62,585,554	1,422,399	613,284–2,503,220	485,190	0.34
NspI-4	48	286,390,596	5,966,471	2,075,002–25,216,938	2,025,629	1.42
NspI-5	48	214,821,584	4,475,450	627,764–6,480,866	953,816	0.43
all F2s	192	664,536,187	3,461,126	613,284–25,216,938	2,906,156	0.59

Table 2.5. Read totals for *X. tropicalis* NspI libraries.

Library	Number of reads	Attributed reads	% attributed
NspI-1	27,905,253	26,463,709	94.8%
NspI-2	62,022,183 pairs	56,867,309 pairs	91.7%
NspI-3	39,951,011 pairs	35,579,594 pairs	89.1%
NspI-4	215,037,446 pairs	143,195,298 pairs	66.6%
NspI-5	121,453,254 pairs	107,410,792 pairs	88.4%

Table 2.6. Phenotypes, barcodes, and read counts for F2 *X. tropicalis* frogs sequenced in NspI libraries.

Library	F2 Frog	Clutch	Color	Sex	Barcode number	Barcode sequence	Read pairs	Total reads
NspI-1	#34	C660	green	female	5	ccagct		1,449,580
NspI-1	#35	C660	green	female	6	ttcaga		1,720,493
NspI-1	#149	C660	gray	female	7	taggaa		1,267,787
NspI-1	#150	C660	gray	female	8	cttgctt		1,126,628
NspI-1	#31	C660	green	male	9	gaacttc		2,118,366
NspI-1	#32	C660	green	male	10	ggaccta		1,541,846
NspI-1	#5	C660	gray	male	11	taggccat		2,278,706
NspI-1	#6	C660	gray	male	12	tgcaagga		2,266,053
NspI-2	#36	C660	green	female	5	ccagct	711,825	1,423,650
NspI-2	#37	C660	green	female	6	ttcaga	876,554	1,753,108
NspI-2	#38	C660	green	female	7	taggaa	992,523	1,985,046
NspI-2	#39	C660	green	female	8	cttgctt	846,430	1,692,860
NspI-2	#40	C660	green	female	9	gaacttc	1,004,134	2,008,268
NspI-2	#41	C660	green	female	10	ggaccta	919,546	1,839,092
NspI-2	#119	C660	green	female	11	taggccat	623,545	1,247,090
NspI-2	#120	C660	green	female	12	tgcaagga	839,990	1,679,980
NspI-2	#121	C660	green	female	13	ctcc	971,090	1,942,180
NspI-2	#122	C660	green	female	14	tgca	929,997	1,859,994
NspI-2	#123	C660	green	female	15	aact	1,063,012	2,126,024
NspI-2	#124	C660	green	female	16	caga	1,108,845	2,217,690
NspI-2	#125	C660	green	female	17	gcgt	1,210,340	2,420,680
NspI-2	#128	C660	green	female	18	gtaa	1,423,872	2,847,744
NspI-2	#151	C660	gray	female	19	cgat	1,109,214	2,218,428
NspI-2	#152	C660	gray	female	20	accgt	1,123,259	2,246,518
NspI-2	#153	C660	gray	female	21	ctagc	1,033,261	2,066,522
NspI-2	#154	C660	gray	female	22	acaaa	1,021,476	2,042,952
NspI-2	#155	C660	gray	female	23	agccc	892,771	1,785,542
NspI-2	#156	C660	gray	female	24	ctgta	1,247,114	2,494,228
NspI-2	#157	C660	gray	female	25	cgctt	867,552	1,735,104
NspI-2	#33	C660	green	male	26	gctta	1,106,904	2,213,808
NspI-2	#132	C660	green	male	27	ggtgt	990,241	1,980,482
NspI-2	#134	C660	green	male	28	tgcga	1,033,321	2,066,642
NspI-2	#135	C660	green	male	29	agtgga	953,444	1,906,888
NspI-2	#137	C660	green	male	30	gctcta	813,852	1,627,704
NspI-2	#138	C660	green	male	31	ccacaa	1,074,783	2,149,566
NspI-2	#139	C660	green	male	32	ggttgt	1,093,353	2,186,706
NspI-2	#140	C660	green	male	33	cttcca	1,163,734	2,327,468
NspI-2	#141	C660	green	male	34	gagata	1,025,747	2,051,494
NspI-2	#142	C660	green	male	35	atgcct	1,076,540	2,153,080
NspI-2	#143	C660	green	male	36	tattttt	887,534	1,775,068
NspI-2	#144	C660	green	male	37	atgaaac	986,724	1,973,448
NspI-2	#147	C660	green	male	38	aaaagtt	589,076	1,178,152
NspI-2	#148	C660	green	male	39	gaattca	1,080,226	2,160,452
NspI-2	#158	C660	gray	male	40	gtcgatt	921,775	1,843,550
NspI-2	#159	C660	gray	male	41	aacgcct	1,032,030	2,064,060
NspI-2	#160	C660	gray	male	42	aatatgc	1,232,570	2,465,140
NspI-2	#161	C660	gray	male	43	acgactac	1,161,196	2,322,392
NspI-2	#162	C660	gray	male	44	tagcagcc	947,603	1,895,206

Table 2.6, continued.

Nspl-2	#163	C660	gray	male	45	tggtacgt	800,152	1,600,304
Nspl-2	#164	C660	gray	male	46	tctcagtc	912,914	1,825,828
Nspl-2	#165	C660	gray	male	47	ccggatat	856,690	1,713,380
Nspl-2	#166	C660	gray	male	48	cgccttat	927,738	1,855,476
Nspl-3	#21	C659	green	female	5	ccagct	385,925	771,850
Nspl-3	#22	C659	green	female	6	ttcaga	630,820	1,261,640
Nspl-3	#23	C659	green	female	7	taggaa	745,701	1,491,402
Nspl-3	#24	C659	green	female	8	cttgctt	581,031	1,162,062
Nspl-3	#25	C659	green	female	9	gaacttc	602,194	1,204,388
Nspl-3	#26	C659	green	female	10	ggaccta	539,259	1,078,518
Nspl-3	#27	C659	green	female	11	taggccat	306,642	613,284
Nspl-3	#61	C659	green	female	12	tgcaagga	398,095	796,190
Nspl-3	#62	C659	green	female	13	ctcc	675,999	1,351,998
Nspl-3	#63	C659	green	female	14	tgca	606,201	1,212,402
Nspl-3	#64	C659	green	female	15	aact	1,023,755	2,047,510
Nspl-3	#105	C659	gray	female	16	caga	770,605	1,541,210
Nspl-3	#106	C659	gray	female	17	gcgt	610,703	1,221,406
Nspl-3	#107	C659	gray	female	18	gtaa	1,285,047	2,570,094
Nspl-3	#108	C659	gray	female	19	cgat	833,166	1,666,332
Nspl-3	#109	C659	gray	female	20	accgt	739,593	1,479,186
Nspl-3	#110	C659	gray	female	21	ctagc	667,810	1,335,620
Nspl-3	#9	C659	green	male	22	acaaa	1,011,397	2,022,794
Nspl-3	#10	C659	green	male	23	agccc	435,291	870,582
Nspl-3	#11	C659	green	male	24	ctgta	967,875	1,935,750
Nspl-3	#12	C659	green	male	25	cgctt	577,767	1,155,534
Nspl-3	#13	C659	green	male	26	gctta	969,710	1,939,420
Nspl-3	#14	C659	green	male	27	ggtgt	680,832	1,361,664
Nspl-3	#15	C659	green	male	28	tgcga	579,737	1,159,474
Nspl-3	#16	C659	green	male	29	agtgga	637,166	1,274,332
Nspl-3	#17	C659	green	male	30	gctcta	723,216	1,446,432
Nspl-3	#18	C659	green	male	31	ccacaa	832,345	1,664,690
Nspl-3	#19	C659	green	male	32	ggttgt	550,127	1,100,254
Nspl-3	#28	C659	gray	male	33	cttcca	800,943	1,601,886
Nspl-3	#29	C659	gray	male	34	gagata	1,099,018	2,198,036
Nspl-3	#30	C659	gray	male	35	atgcct	735,925	1,471,850
Nspl-3	#111	C659	gray	male	36	tattttt	1,251,610	2,503,220
Nspl-3	#112	C659	gray	male	37	atgaaac	1,002,278	2,004,556
Nspl-3	#113	C659	gray	male	38	aaaagtt	822,017	1,644,034
Nspl-3	#114	C659	gray	male	39	gaattca	880,970	1,761,940
Nspl-3	#115	C659	gray	male	40	gtcgatt	479,486	958,972
Nspl-3	#116	C659	gray	male	41	aacgcct	535,692	1,071,384
Nspl-3	#126	C659	green	female	42	aatatgc	1,185,653	2,371,306
Nspl-3	#127	C659	green	female	43	acgactac	763,882	1,527,764
Nspl-3	#129	C659	green	female	44	tagcagcc	354,539	709,078
Nspl-3	#130	C659	green	female	45	tggtacgt	654,286	1,308,572
Nspl-3	#133	C659	green	male	46	tctcagtc	582,968	1,165,936
Nspl-3	#145	C659	green	male	47	ccggatat	386,533	773,066
Nspl-3	#146	C659	green	male	48	cgccttat	388,968	777,936
Nspl-4	#286	C1029	tadpole		1	acta	4,172,869	8,345,738
Nspl-4	#287	C1029	tadpole		2	tcacc	1,786,418	3,572,836
Nspl-4	#288	C1029	tadpole		3	ttctc	3,035,871	6,071,742
Nspl-4	#289	C1029	tadpole		4	gtatt	2,887,265	5,774,530

Table 2.6, continued.

Nspl-4	#290	C1029	tadpole	5	ccagct	1,942,378	3,884,756
Nspl-4	#293	C1029	tadpole	6	ttcaga	3,307,981	6,615,962
Nspl-4	#294	C1029	tadpole	7	taggaa	4,118,991	8,237,982
Nspl-4	#295	C1029	tadpole	8	cttgctt	3,225,831	6,451,662
Nspl-4	#296	C1029	tadpole	9	gaacttc	2,437,865	4,875,730
Nspl-4	#297	C1029	tadpole	10	ggaccta	2,010,961	4,021,922
Nspl-4	#298	C1029	tadpole	11	taggccat	1,253,413	2,506,826
Nspl-4	#299	C1029	tadpole	12	tgcaagga	1,037,501	2,075,002
Nspl-4	#300	C1029	tadpole	13	ctcc	2,270,640	4,541,280
Nspl-4	#301	C1029	tadpole	14	tgca	1,962,978	3,925,956
Nspl-4	#302	C1029	tadpole	15	aact	4,147,464	8,294,928
Nspl-4	#303	C1029	tadpole	16	caga	5,319,921	10,639,842
Nspl-4	#304	C1029	tadpole	17	gcgt	2,816,257	5,632,514
Nspl-4	#305	C1029	tadpole	18	gtaa	7,287,028	14,574,056
Nspl-4	#306	C1029	tadpole	19	cgat	2,770,068	5,540,136
Nspl-4	#307	C1029	tadpole	20	accgt	4,766,249	9,532,498
Nspl-4	#308	C1029	tadpole	21	ctagc	2,093,579	4,187,158
Nspl-4	#309	C1029	tadpole	22	acaaa	4,853,787	9,707,574
Nspl-4	#310	C1029	tadpole	23	agccc	1,140,928	2,281,856
Nspl-4	#311	C1029	tadpole	24	ctgta	5,394,088	10,788,176
Nspl-4	#313	C1029	tadpole	25	cgctt	1,274,482	2,548,964
Nspl-4	#314	C1029	tadpole	26	gctta	3,705,724	7,411,448
Nspl-4	#315	C1029	tadpole	27	ggtgt	2,462,871	4,925,742
Nspl-4	#316	C1029	tadpole	28	tgcga	1,728,977	3,457,954
Nspl-4	#318	C1029	tadpole	29	agtgga	2,227,815	4,455,630
Nspl-4	#319	C1029	tadpole	30	gctcta	2,094,089	4,188,178
Nspl-4	#321	C1029	tadpole	31	ccacaa	2,866,072	5,732,144
Nspl-4	#322	C1029	tadpole	32	ggttgt	1,888,476	3,776,952
Nspl-4	#324	C1029	tadpole	33	cttcca	1,254,586	2,509,172
Nspl-4	#327	C1029	tadpole	34	gagata	2,095,449	4,190,898
Nspl-4	#328	C1029	tadpole	35	atgcct	1,107,230	2,214,460
Nspl-4	#329	C1029	tadpole	36	tattttt	4,399,023	8,798,046
Nspl-4	#330	C1029	tadpole	37	atgaaac	4,732,101	9,464,202
Nspl-4	#331	C1029	tadpole	38	aaaagtt	12,608,469	25,216,938
Nspl-4	#332	C1029	tadpole	39	gaattca	1,140,732	2,281,464
Nspl-4	#333	C1029	tadpole	40	gtcgatt	2,647,258	5,294,516
Nspl-4	#334	C1029	tadpole	41	aacgcct	1,746,807	3,493,614
Nspl-4	#336	C1029	tadpole	42	aatatgc	5,916,728	11,833,456
Nspl-4	#337	C1029	tadpole	43	acgactac	2,564,200	5,128,400
Nspl-4	#338	C1029	tadpole	44	tagcagcc	1,533,024	3,066,048
Nspl-4	#339	C1029	tadpole	45	tggtacgt	2,650,907	5,301,814
Nspl-4	#340	C1029	tadpole	46	tctcagtc	1,826,854	3,653,708
Nspl-4	#341	C1029	tadpole	47	ccggatat	1,635,837	3,271,674
Nspl-4	#342	C1029	tadpole	48	cgccttat	1,047,256	2,094,512
Nspl-5	#292	C1029	tadpole	1	acta	1,710,276	3,420,552
Nspl-5	#343	C1029	tadpole	2	tcacc	2,091,404	4,182,808
Nspl-5	#344	C1029	tadpole	3	ttctc	1,661,656	3,323,312
Nspl-5	#345	C1029	tadpole	4	gtatt	1,781,582	3,563,164
Nspl-5	#346	C1029	tadpole	5	ccagct	2,321,448	4,642,896
Nspl-5	#347	C1029	tadpole	6	ttcaga	1,858,822	3,717,644
Nspl-5	#348	C1029	tadpole	7	taggaa	2,248,314	4,496,628
Nspl-5	#349	C1029	tadpole	8	cttgctt	1,363,366	2,726,732

Table 2.6, continued.

Nspl-5	#351	C1029	tadpole	9	gaacttc	2,564,310	5,128,620
Nspl-5	#352	C1029	tadpole	10	ggaccta	1,536,598	3,073,196
Nspl-5	#353	C1029	tadpole	11	taggccat	2,105,118	4,210,236
Nspl-5	#354	C1029	tadpole	12	tgcaagga	1,540,826	3,081,652
Nspl-5	#356	C1029	tadpole	13	ctcc	1,707,280	3,414,560
Nspl-5	#357	C1029	tadpole	14	tgca	1,428,904	2,857,808
Nspl-5	#358	C1029	tadpole	15	aact	1,571,954	3,143,908
Nspl-5	#359	C1029	tadpole	16	caga	2,037,038	4,074,076
Nspl-5	#360	C1029	tadpole	17	gcgt	2,365,682	4,731,364
Nspl-5	#361	C1029	tadpole	18	gtaa	3,021,322	6,042,644
Nspl-5	#362	C1029	tadpole	19	cgat	4,580,028	9,160,056
Nspl-5	#363	C1029	tadpole	20	accgt	2,281,180	4,562,360
Nspl-5	#366	C1029	tadpole	21	ctagc	2,521,052	5,042,104
Nspl-5	#367	C1029	tadpole	22	acaaa	2,052,724	4,105,448
Nspl-5	#368	C1029	tadpole	23	agccc	2,147,530	4,295,060
Nspl-5	#369	C1029	tadpole	24	ctgta	2,757,898	5,515,796
Nspl-5	#370	C1029	tadpole	25	cgctt	1,477,472	2,954,944
Nspl-5	#371	C1029	tadpole	26	gctta	2,834,886	5,669,772
Nspl-5	#372	C1029	tadpole	27	ggtgt	2,623,712	5,247,424
Nspl-5	#373	C1029	tadpole	28	tgcga	1,906,884	3,813,768
Nspl-5	#374	C1029	tadpole	29	agtgga	4,013,372	8,026,744
Nspl-5	#375	C1029	tadpole	30	gctcta	3,225,562	6,451,124
Nspl-5	#377	C1029	tadpole	31	ccacaa	2,793,030	5,586,060
Nspl-5	#378	C1029	tadpole	32	ggttgt	2,235,782	4,471,564
Nspl-5	#379	C1029	tadpole	33	cttcca	6,480,866	12,961,732
Nspl-5	#380	C1029	tadpole	34	gagata	2,597,974	5,195,948
Nspl-5	#381	C1029	tadpole	35	atgcct	2,771,794	5,543,588
Nspl-5	#382	C1029	tadpole	36	tattttt	1,017,242	2,034,484
Nspl-5	#384	C1029	tadpole	37	atgaaac	1,716,676	3,433,352
Nspl-5	#386	C1029	tadpole	38	aaaagtt	2,033,644	4,067,288
Nspl-5	#387	C1029	tadpole	39	gaattca	2,476,620	4,953,240
Nspl-5	#388	C1029	tadpole	40	gtcgatt	1,916,988	3,833,976
Nspl-5	#390	C1029	tadpole	41	aacgcct	2,233,142	4,466,284
Nspl-5	#391	C1029	tadpole	42	aatatgc	1,381,970	2,763,940
Nspl-5	#393	C1029	tadpole	43	acgactac	2,305,470	4,610,940
Nspl-5	#394	C1029	tadpole	44	tagcagcc	1,564,814	3,129,628
Nspl-5	#399	C1029	tadpole	45	tggtacgt	2,943,170	5,886,340
Nspl-5	#403	C1029	tadpole	46	tctcagtc	627,764	1,255,528
Nspl-5	#406	C1029	tadpole	47	ccggatat	960,758	1,921,516
Nspl-5	#408	C1029	tadpole	48	cgccttat	2,014,888	4,029,776

IIIc. The sex determining locus and the *gray* gene

As noted above, the sex and color (green or gray) was recorded for 96 of the F2s we sequenced (Table 2.6). Therese used the approximately 16,000 markers that were genotyped in at least half of the F2s to calculate correlations for the sex-determining and *gray* loci. I will discuss the sex locus in Chapter Three. The places in the v. 7 assembly that showed a correlation of greater than 0.4 with the *gray* locus, and the genes that are nearest to those places, are listed in Table 2.7.

As described by Lane et al., *gray* is a recessive background mutation that emerged during the inbreeding of the ICB line of *X. tropicalis*. The *gray* phenotype is caused by a single locus: it displays a Mendelian inheritance pattern and matings between *gray* mutants have produced 100% gray offspring for several generations. The *gray* phenotype results from defects in both number and morphology of the yellow xanthophores as well as the iridescent iridophores in the adult dorsal skin.

A number of pigmentation mutants have been reported in zebrafish, and I investigated whether any of these might be good candidates for *gray*. Of particular note, the mutations *pfeffer* and *salz* cause a loss of xanthophores (Odenthal et al., 1996). *salz* does not appear to have been mapped. Although the gene causing *pfeffer* has not been identified, based on its location in linkage group 14 (zfin.org) and synteny with the *X. tropicalis* genome, it does not seem that *pfeffer* lies on an *X. tropicalis* scaffold with correlation to the *gray* gene. Parichy and Turner reported that the gene *fms*, now called *csf1r*, is required for xanthophore development (Parichy and Turner, 2003). In *X. tropicalis*, *csf1r* is not on a scaffold with correlation to the *gray* gene, nor are any other annotated *csf* ligand or receptor genes.

Another well-known zebrafish pigmentation mutant is *golden*, which affects melanin pigmentation and is caused by a mutation in the cation exchanger *slc24a5*. The *X. tropicalis* ortholog of *slc24a5* (confirmed by synteny) lies on v.7 scaffold 3; no correlation was found between the *gray* locus and scaffold 3. Two slc genes, *slc47a2* and *slc38a5*, lie in the *gray*-correlated regions (Table 2.7), but as there are 351 *slc* genes annotated in the *X. tropicalis* genome (xenbase.org) the observation is suggestive but may not be significant.

The Rokhsar lab is currently working on a new assembly for *X. tropicalis* by linking together the supercontigs from version 7 using the SNP map Therese generated from our GBS data. It is possible that some or all of the regions of correlation to *gray* will be united in the new assembly, which might result in one region showing very strong correlation to the *gray* mutation. At that point, the next step would be to utilize three- and four-allele SNP sites (which were excluded from the original set) for finer mapping, and then prioritize and investigate candidates for the *gray* gene.

Table 2.7. Markers with correlation to the *gray* mutation.

v.7 scaffold	Position	Correlation	Closest genes
2	37,921,685	0.50	at 37,600 kb: zpld1; between 38,200 kb and 38,270 kb: FAM172-like, txn14b, rg9mtd1, pcnp, zbtb11, rpp21, clqbp
2	37,921,695	0.52	
2	37,921,706	0.52	
2	37,921,963	0.44	
8b	1,461,810	0.62	abca2, oct60
8b	1,967,803	0.75	lrsam1, anapc2, ssna1, slc47a2
8b	1,968,653	0.72	
266	60,997	0.43	at 45 kb: hypothetical protein LOC100497513 (no domains predicted by SMART); at 79–110 kb: ddr2
266	61,426	0.59	
979	12,238	0.64	slc38a5 (7 kb–17 kb; only gene on scaffold)
979	12,271	0.69	
979	12,627	0.61	
1757	4,164		synaptonemal complex central element protein 1-like (based on EST cluster; no v7.2 annotation)

IV. Genotyping by sequencing in *X. laevis*

X. laevis is allotetraploid, and up to half of the genes retain two homeologs, also often referred to as alloalleles (Hellsten et al., 2007), which presents challenges for genome assembly. I thus sought to adapt my GBS protocol for *X. laevis* to facilitate the construction of a SNP map that could be used to link the small scaffolds together and thereby contribute to improving the genome assembly. My collaborators in the Rokhsar lab and I wanted to attain denser coverage of the *X. laevis* genome than we had with *X. tropicalis* in order to ensure that as many scaffolds as possible contained at least one marker. The Rokhsar lab anticipates being able to use synteny with the *X. tropicalis* genome for assembly, but to reliably separate the homeologs and designate each as an A or B gene (on an A or B chromosome) they need a dense map, though not necessarily a high resolution one. We chose to use the restriction enzyme NlaIII because it leaves the same overhang as NspI, allowing me to use the same adapters as for the *X. tropicalis* map. However, NlaIII cuts more frequently than NspI due to its four-base-pair recognition site, and thus should yield a larger number of fragments and a greater number of informative markers [Figure 2.1(1)]. By selecting the fraction of ligated DNA fragments from 400–800 bp, I expect to have sampled a large proportion of the genome [Figure 2.1(4)].

I am sequencing DNA from *X. laevis* frogs from crosses between the inbred F and J strains, using DNA from blood or tadpoles sent to us by the Robert lab at the University of Rochester. The J strain, originally from South Africa, has been inbred for approximately 30 generations (D. Kelly and M. Taira, personal communications). The F strain, also originally from South Africa, has been inbred since the 1970's but with occasional outbreeding (J. Robert, personal communication). The approximate relationships between the *X. laevis* frogs we are using for GBS are shown in Figure 2.7.

The frog used to construct the *X. laevis* genome was from the J strain. Since this strain is so inbred, we will use the genome assembly, and the sequence reads used to construct it, to assess the J genotype. For the F strain female, we constructed a sheared/shotgun library using the Illumina Truseq kit (Table 2.1), which we sequenced as both 100 bp and 150 bp paired-end reads.

We received lysed F/J x F/J (F2) and F/J x J (backcross) tadpoles from the Robert lab, and DNA was isolated as described below in Additional materials and methods. I constructed a pooled multiplex library for each group: the backcross library comprised 39 individuals and the F2 library comprised 29 individuals. Many of the F2 tadpole DNAs had appeared degraded when run on a gel, so I used only those that had a visible high-molecular-weight band although these also contained some degraded DNA.

I used DNA from two F/J hybrids to represent the F1 generation. These individuals are closely related to the parents of the F2 tadpoles, or may have been their actual parents (Figure 2.7). DNA was isolated from lysed blood provided by the Robert lab as described below in Additional materials and methods, and individual libraries were made using these DNAs, in parallel with the F2 and backcross libraries. The only major differences in the library preparation protocol from that used for the *X. tropicalis* libraries were the change in

restriction enzyme, the broader size selection, and reduction in PCR cycles to 10. Specific library preparation parameters are listed in Table 2.1, and the Bioanalyzer results are in Figure 2.8. The F/J, backcross, and F2 libraries were sequenced as 150 bp paired-end reads.

Although the two F/J hybrids were originally intended to be combined in with the F2 or backcross libraries, we decided to run them together on their own lane in order to fill a flow cell. The poor base balance of their barcodes, which share the same base at position 2, however, probably caused the sequencing failure of read 1 in two separate runs (Table 2.8, Figure 2.9). Since read 2 sequenced reasonably well, though, my collaborators and I believe we have enough data from the F1 frogs to proceed.

The F/J x J library sequenced well except for an as-yet-unexplained drop in quality scores around 100 bp, but the F/J x F/J library failed (Figure 2.10), presumably due to the poor DNA quality noted above.

As a test of whether the NspI method will be effective for *X. laevis*, my collaborator Therese Mitros performed an alignment of the reads from one F/J x J individual to *X. laevis* genome assembly version 5, and approximately 86% of the reads aligned successfully. Based on this test, we decided to construct another library comprising the 11 remaining backcross tadpole DNAs as well as a newer set of 30 F/J x F/J tadpole DNAs that appeared to be of much better quality than those used previously. In the hopes of reducing read count heterogeneity by more precisely using the same amount of DNA for each individual, I added a second Picogreen quantitation step after an initial dilution to approximately 30 ng/μL. Bioanalyzer results for this library (NlaIII-B) are in Figure 2.8.

All individual *X. laevis* DNAs used in multiplex libraries that sequenced successfully are listed in Table 2.8 with their assigned barcodes. Beyond the alignment test and the quality graphs given in Figure 2.9 and Figure 2.10, the data from these libraries have not yet been analyzed.

The work presented in this chapter demonstrates the transformative power of nextgen sequencing technology on our approach to genetics. My approach for reduced-representation multiplex nextgen sequencing facilitates the genotyping of many individuals at low cost, allowing for the relatively quick construction of genetic maps, mapping of genetic loci, and genome assembly. Given the rapid pace of improvements in nextgen sequencing technology, we will be able to genotype ever larger numbers of individuals at lower cost, which will further increase the resolution and precision of genetic maps thus generated.

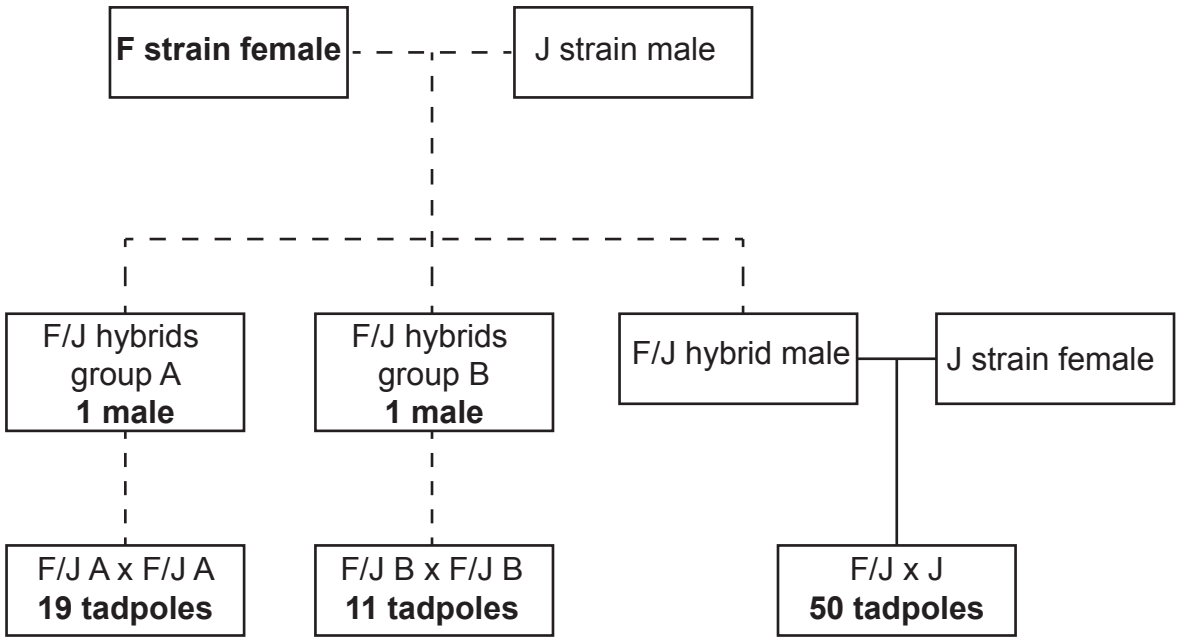
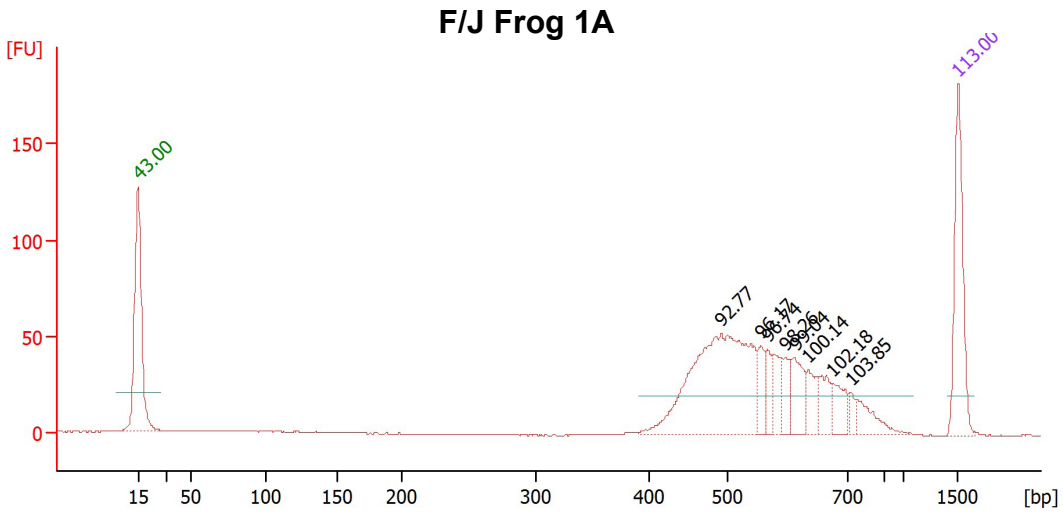


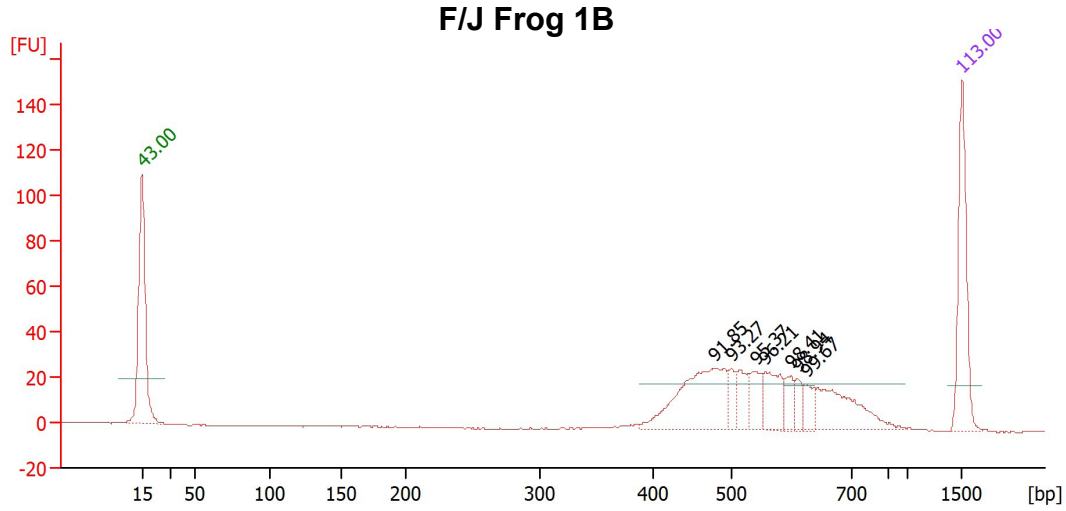
Figure 2.7. Genealogy of *X. laevis* frogs genotyped by sequencing. We sequenced *X. laevis* frogs descended from F and J strains. The dotted lines indicate that the individual we sequenced may or may not have been the parent of the next generation. Sequenced individuals are indicated in bold.

Figure 2.8. Bioanalyzer results for *X. laevis* NlaIII libraries.



Peak table :

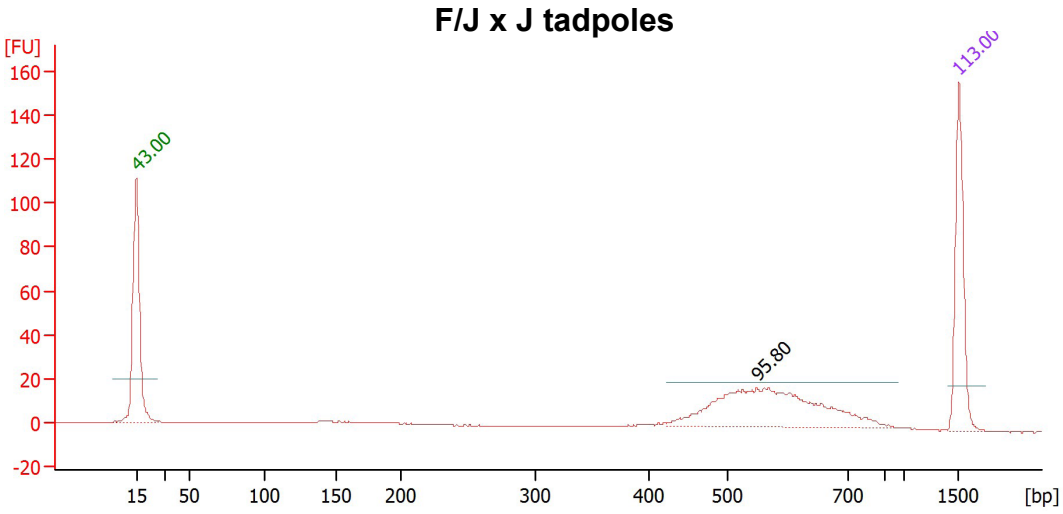
Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		492	10.68	32.9	
10	▶	1,500	2.10	2.1	Upper Marker



Peak table :

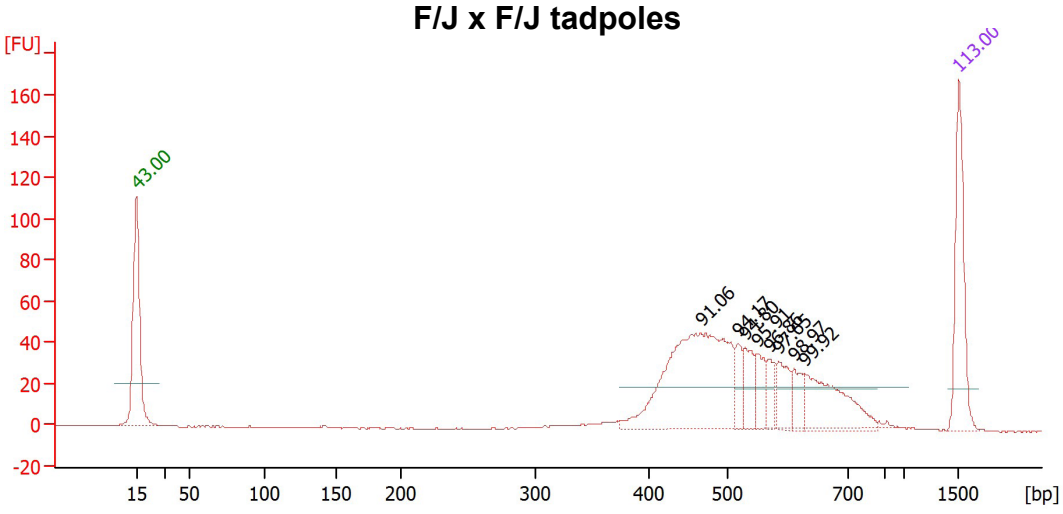
Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		479	7.77	24.6	
9	▶	1,500	2.10	2.1	Upper Marker

Figure 2.8, continued.



Peak table :

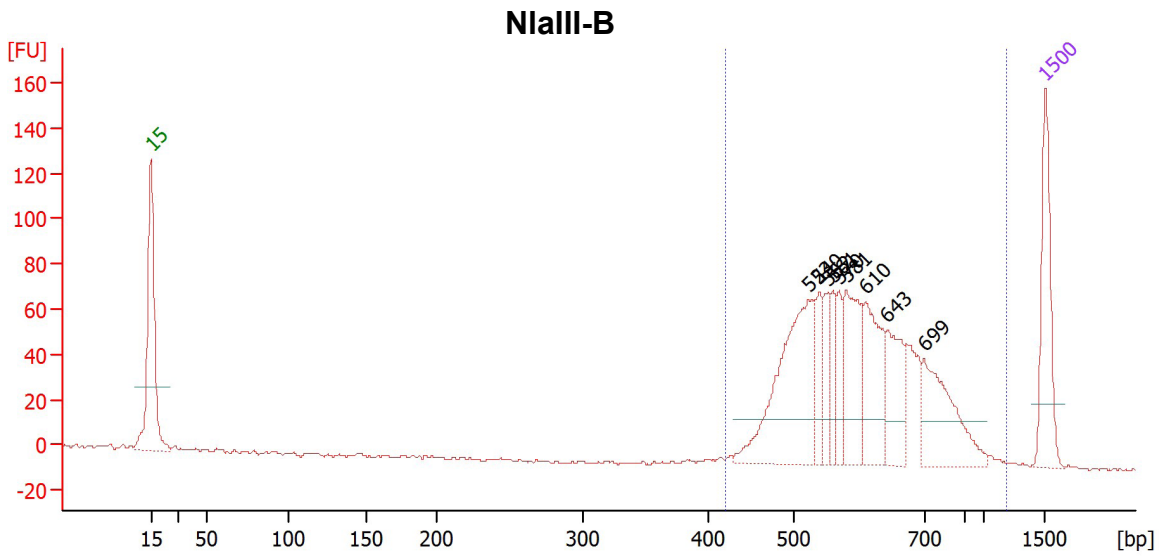
Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		550	3.61	10.0	
3	▶	1,500	2.10	2.1	Upper Marker



Peak table :

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		467	10.95	35.6	
10	▶	1,500	2.10	2.1	Upper Marker

Figure 2.8, continued.



Peak table :

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	523	4.38	12.7	
3	540	0.82	2.3	
4	552	0.69	1.9	
5	561	0.76	2.0	
6	570	0.68	1.8	
7	581	1.93	5.0	
8	610	2.23	5.5	
9	643	1.56	3.7	
10	699	2.35	5.1	
11	1,500	2.10	2.1	Upper Marker

Region table :

From [bp]	To [bp]	Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [ng/µl]	Color
421	1,175	895.0	99	606	16.6	16.15	

Table 2.8. *X. laevis* individuals sequenced in NlaIII libraries.

Library	Frog	Barcode number	Barcode Sequence
F/J males	F/J Frog1A	1	acta
F/J males	F/J Frog1B	2	tcacc
F/JxJ	F/JxJ 1	5	ccagct
F/JxJ	F/JxJ 2	6	ttcaga
F/JxJ	F/JxJ 3	7	taggaa
F/JxJ	F/JxJ 4	8	cttgctt
F/JxJ	F/JxJ 5	9	gaacttc
F/JxJ	F/JxJ 6	10	ggaccta
F/JxJ	F/JxJ 7	11	taggccat
F/JxJ	F/JxJ 8	12	tgcaagga
F/JxJ	F/JxJ 9	13	ctcc
F/JxJ	F/JxJ 10	14	tgca
F/JxJ	F/JxJ 11	15	aact
F/JxJ	F/JxJ 12	16	caga
F/JxJ	F/JxJ 13	17	gcgt
F/JxJ	F/JxJ 14	18	gtaa
F/JxJ	F/JxJ 15	19	cgat
F/JxJ	F/JxJ 16	20	accgt
F/JxJ	F/JxJ 17	21	ctagc
F/JxJ	F/JxJ 18	22	acaaa
F/JxJ	F/JxJ 19	23	agccc
F/JxJ	F/JxJ 20	24	ctgta
F/JxJ	F/JxJ 21	25	cgctt
F/JxJ	F/JxJ 22	26	gctta
F/JxJ	F/JxJ 23	27	ggtgt
F/JxJ	F/JxJ 24	28	tgcga
F/JxJ	F/JxJ 25	29	agtgga
F/JxJ	F/JxJ 26	30	gctcta
F/JxJ	F/JxJ 27	31	ccacaa
F/JxJ	F/JxJ 28	32	ggttgt
F/JxJ	F/JxJ 29	33	cttcca
F/JxJ	F/JxJ 30	34	gagata
F/JxJ	F/JxJ 31	35	atgcct
F/JxJ	F/JxJ 32	36	tattttt
F/JxJ	F/JxJ 33	37	atgaaac
F/JxJ	F/JxJ 34	38	aaaagtt
F/JxJ	F/JxJ 35	39	gaattca
F/JxJ	F/JxJ 36	40	gtcgatt
F/JxJ	F/JxJ 37	41	aacgcct
F/JxJ	F/JxJ 38	42	aatatgc
F/JxJ	F/JxJ 39	43	acgactac
NlaIII-B	F/JxJ 40	8	cttgctt

Table 2.8, continued.

NlaIII-B	F/JxJ 41	9	gaacttc
NlaIII-B	F/JxJ 42	10	ggaccta
NlaIII-B	F/JxJ 43	11	taggccat
NlaIII-B	F/JxJ 44	12	tgcaagga
NlaIII-B	F/JxJ 45	13	ctcc
NlaIII-B	F/JxJ 46	14	tgca
NlaIII-B	F/JxJ 47	15	aact
NlaIII-B	F/JxJ 48	16	caga
NlaIII-B	F/JxJ 49	17	gcgt
NlaIII-B	F/JxJ 50	18	gtaa
NlaIII-B	F/JxF/J A26	19	cgat
NlaIII-B	F/JxF/J A27	20	accgt
NlaIII-B	F/JxF/J A28	21	ctagc
NlaIII-B	F/JxF/J A29	22	acaaa
NlaIII-B	F/JxF/J A30	23	agccc
NlaIII-B	F/JxF/J A31	24	ctgta
NlaIII-B	F/JxF/J A32	25	cgctt
NlaIII-B	F/JxF/J A33	26	gctta
NlaIII-B	F/JxF/J A34	27	ggtgt
NlaIII-B	F/JxF/J A35	28	tgcga
NlaIII-B	F/JxF/J A36	29	agtgga
NlaIII-B	F/JxF/J A37	30	gctcta
NlaIII-B	F/JxF/J A38	31	ccacaa
NlaIII-B	F/JxF/J A39	32	ggttgt
NlaIII-B	F/JxF/J A41	33	cttcca
NlaIII-B	F/JxF/J A42	34	gagata
NlaIII-B	F/JxF/J A43	35	atgcct
NlaIII-B	F/JxF/J A44	36	tattttt
NlaIII-B	F/JxF/J A45	37	atgaaac
NlaIII-B	F/JxF/J B42	38	aaaagtt
NlaIII-B	F/JxF/J B43	39	gaattca
NlaIII-B	F/JxF/J B47	40	gtcgatt
NlaIII-B	F/JxF/J B48	41	aacgcct
NlaIII-B	F/JxF/J B49	42	aatatgc
NlaIII-B	F/JxF/J B50	43	acgactac
NlaIII-B	F/JxF/J B51	44	tagcagcc
NlaIII-B	F/JxF/J B52	45	tggtacgt
NlaIII-B	F/JxF/J B53	46	tctcagtc
NlaIII-B	F/JxF/J B54	47	ccggatat
NlaIII-B	F/JxF/J B55	48	cgccttat

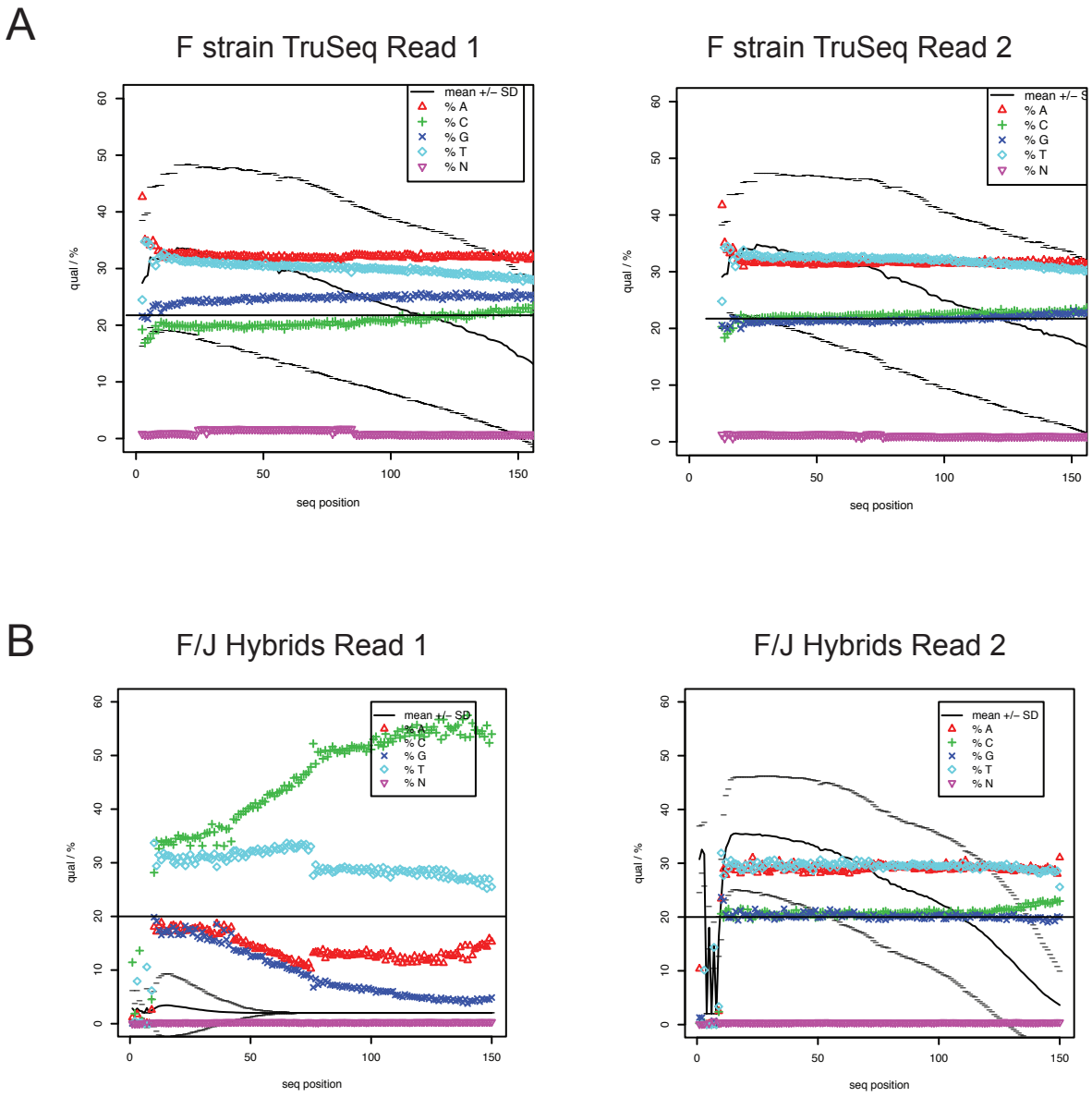


Figure 2.9. Quality score and base percentage graphs for *X. laevis* F strain and F/J hybrid libraries. Graphs generated by T. Mitros, based on a sample of 200,000 reads each. (A) F strain TruSeq library 150 bp paired-end results. The reason for the notable gap between A and T, and G and C, percentages in the forward read is unknown. 100 bp paired-end results were similar; average quality scores remained greater than 20 (95% confidence) past 90 bp. (B) The F/J hybrid library, which consists of two individuals combined onto one lane, was sequenced twice as 150 bp paired-end reads with similar results both times. The failure of read 1 and the drop in quality scores at the beginning of read 2 may be due to the two barcodes having the same base at position 2 of their barcodes.

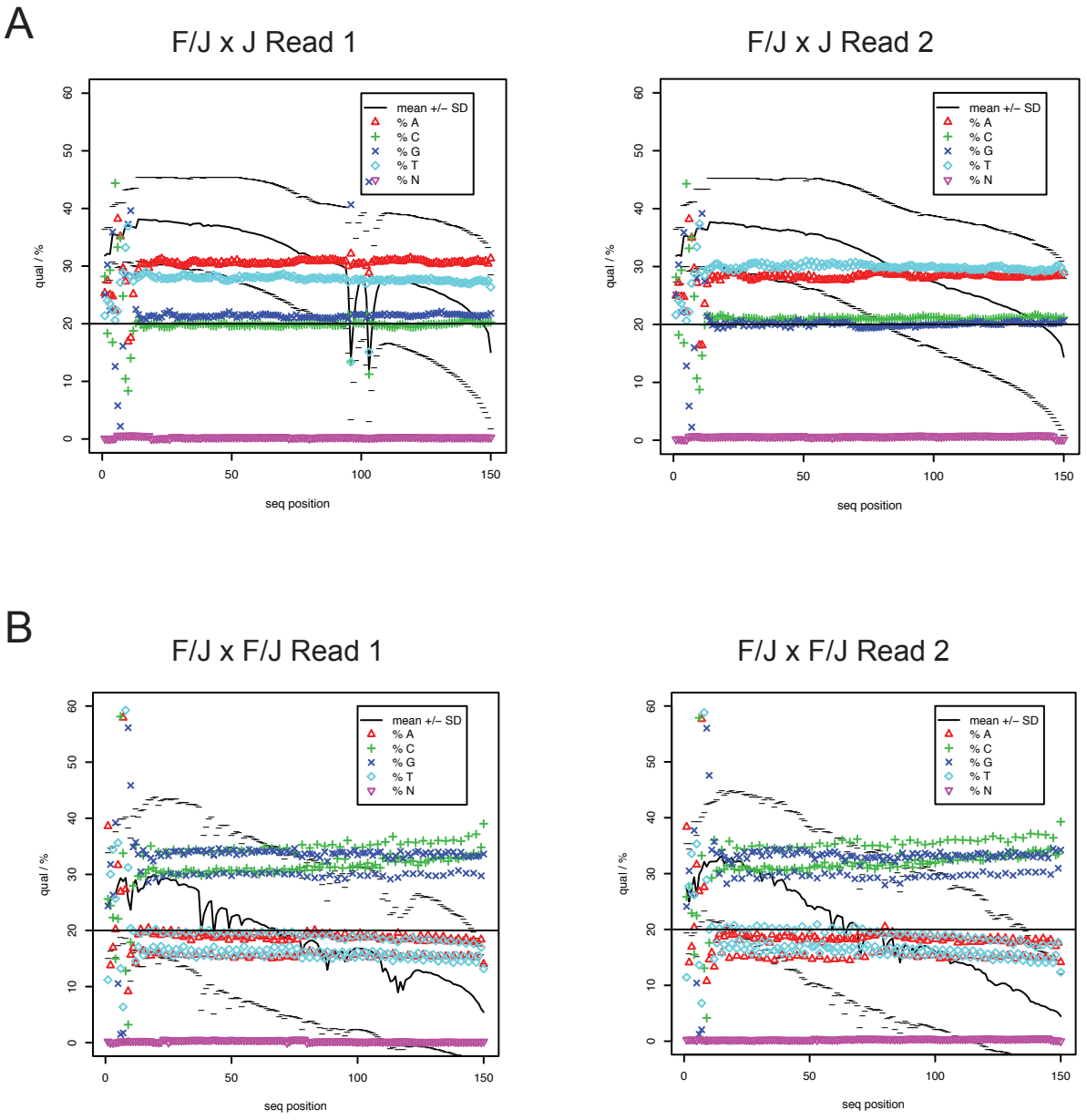


Figure 2.10. Quality score and base percentage graphs for *X. laevis* F/J x J and F/J x F/J libraries. Graphs generated by T. Mitros, based on a sample of 200,000 reads each. (A) F/J x J library 150 bp paired-end results. The heterogeneity in basecall percentages at the beginning of the reads is due to barcoding, but the reason for the drop in quality scores around 100 bp in read 1 is unknown. (B) F/J x F/J library 150 bp paired-end results. This library failed: quality scores drop below 20 halfway through the reads, and basecalls appear periodic. We attribute the failure of this library to poor DNA quality.

V. Additional materials and methods

Specific preparation parameters for each Illumina library are given in Table 2.1.

X. tropicalis P0 and F1; *X. laevis* F and F/J hybrid DNAs; and all tadpole DNAs were isolated from toe clips (*X. tropicalis* P0s and F1s), blood (*X. laevis* F and F/J), or tadpoles lysed for 16–120 hours at 55°C in lysis buffer (50 mM Tris pH 7.5, 62.5 mM NaCl, 5 mM EDTA, 0.5% SDS, 5% Chelex resin [Bio-Rad]) supplemented with Proteinase K to 250–500 μg/mL, then mixed and spun down. The supernatant was removed, Phenol/Chloroform extracted, and resuspended in TE, Tris, or water.

The *X. tropicalis* adult F2 DNAs were isolated from red blood cells essentially as described at http://tropicalis.berkeley.edu/home/genetic_techniques/genomicDNA.html and resuspended in TE pH 8. Most of these DNAs were further diluted 1:5 in water before use.

DNA quantitation was performed with Picogreen reagent (Invitrogen). All DNAs were stored at 4°C.

PCR program for multiplex libraries with custom barcoded adapters (Elshire et al., 2011):
98°C for 30 sec
10–15 cycles of:
    98°C for 10 sec
    65°C for 30 sec
    72°C for 30 sec
72°C for 5 min

The size distribution and approximate concentration of Illumina libraries was assessed on an Agilent 2100 Bioanalyzer using the DNA 7500, DNA 1000, or High Sensitivity assay in the Functional Genomics Laboratory at UC Berkeley. Sequencing was performed by the Vincent Coates Genomic Sequencing Laboratory at UC Berkeley.

VI. GBS protocol for tadpole DNAs

Anneal adapters:
• Dilute oligos to 200 μM in TE, pipet up and down to mix well.

• In a 96-well plate, for each adapter, prepare annealing mix:
15 μL "F" oligo
15 μL "R" oligo
30 μL 2X annealing buffer (100 mM NaCl, 20 mM Tris-Cl)[1]
→In a thermocycler, anneal adapters using the following program:
95°C 4 min, ramp to 25°C by 0.1°C/sec, 25°C 30 min, cool to 4°C

• To check the annealing, run 1 μL of each annealed adapter next to 1 μL of unannealed adapter on a 4% agarose gel. The annealed adapter should appear larger.

• Dilute adapters to 10 μM in water and quantitate with Picogreen.

• Make 2.5 μM dilutions of adapters in water based on the calculated molecular weight of the double-stranded portion of each adapter.[2]

Perform digest:
• Quantitate DNA with Picogreen reagent[3]

• Make 30 ng/μL dilutions of each DNA in 50 μL; quantitate diluted DNAs using Picogreen reagent.

• For each sample:

100 ng	diluted DNA
0.2 μL	100X BSA
2 μL	10X restriction enzyme buffer
2U	restriction enzyme
to 20 μL	H_20

→Incubate 4 hours 37°C, 20 min 65°C (to heat inactivate enzyme), cool to 4°C or freeze.

Ligation:
• Bring 2.5 μM annealed adapters to room temperature, spin down
• Bring digests to room temperature, spin down if needed

• Add 2 μL 2.5 μM annealed adapter to each well, using a different barcoded adapter for each individual.

---

[1] Although some salt is needed to raise the Tm of the annealed adapters, it may not be critical for it to be this particular buffer.
[2] Picogreen only quantitates double-stranded DNA.
[3] Picogreen does not accurately quantitate DNA of concentration > 200ng/μL.

• Prepare ligation cocktail at room temperature. For each sample:

4 μL	10X T4 ligase buffer (New England Biolabs)
1 μL	T4 ligase (New England Biolabs)
13 μL	H_20

• Add 18 μL ligation mix to each well, pipet up and down to mix.
→ Incubate 2 hours 22°C, 30 min 65°C, cool to 4°C or freeze

Pooling and cleanup:
• If preparing 48 samples, pipet 8.33 μL from each ligation into 2000 μL Qiagen buffer PB. Pipet up and down when adding each sample, to ensure that the same amount is added from each sample.

• Clean using MinElute kit, per kit instructions, elute with 25 μL buffer EB.

Size selection:
• Run pooled ligations on a 2% agarose gel at ~140V.

• Excise a gel slice at the desired size range and gel extract using MinElute kit per kit instructions, melting the gel at room temperature. Elute with 22 μL buffer EB.

• Repeat size selection, elute with 22 μL buffer EB.

PCR:

10 μL	size-selected pooled DNA
4 μL	primer PE1 (0.2 uM final)
4 μL	primer PE2 (0.2 uM final)[4]
2 μL	10 mM dNTPs (0.4mM each final)
1 μL	Phusion polymerase (New England Biolabs)
10 μL	5X HF buffer (New England Biolabs)
19 μL	H_20
50 μL	total

Program:
98°C, 30 sec
10 cycles:
    98°C, 10 sec
    65°C, 30 sec
    72°C, 30 sec
72°C, 5 min

[4] PE1 sequence = 5'CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCT GCTGAACCGCTCTTCCGATCT 3'
PE2 sequence = 5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACA CGACGCTCTTCCGATCT 3'

• Clean PCR with 0.7 volumes AMPure XP beads.

Chapter Three: The mode and mechanism of sex determination in *X. tropicalis*

I. Introduction

Sex determination is a rapidly evolving trait, and vertebrates exhibit a wide variety of mechanisms of sex determination, including genetic and environmentally-determined modes (Ezaz et al., 2006). There are two classes of genetic modes of sex determination: XX/XY and WZ/ZZ. In an XX/XY system, males are the heterogametic sex, whereas in a WZ/ZZ system, the female is the heterogametic sex. Either of these modes can be affiliated with various numbers of sex chromosomes, which may or may not be evident cytologically (Ezaz et al., 2006).

As with most amphibians (Hayes, 1998), *Xenopus* don't have sex chromosomes that differ cytologically from autosomes (Yoshimoto et al., 2008). We know that in *X. laevis* and *X. tropicalis,* sex is not determined environmentally, because clutches of frogs raised together in the same facility yield males and females in approximately equal numbers. Furthermore, Hayes contends that environment does not affect amphibian sex determination in the wild, because they don't encounter the temperatures that have been shown in the lab to affect sex determination (Hayes, 1998).

Chang and Witschi proved that *X. laevis* uses a WZ/ZZ mode of sex determination: out of a group of female frogs that had been exposed to estrogen during their development, approximately half gave all male offspring (as in Figure 3.1A, left side)(CHANG and WITSCHI, 1956). Recently, Yoshimoto and colleagues demonstrated that a gene called *DM-W*, present only in females, is the sex-determining gene in *X. laevis* (Yoshimoto et al., 2010). DM-W is related to the transcription factor DMRT1, which is required for testis formation, and it has the DNA-binding DM domain, but no transactivating domain. Thus, it acts as a dominant negative, preventing testis development and thereby driving the bipotential gonad toward the female ovary fate. The mode and mechanism of sex determination in *X. tropicalis* are not known: *DM-W* has not been found in the *X. tropicalis* genome, however, and Bewick et al. report that *DM-W* evolved after the divergence of *X. laevis* and *X. tropicalis* (Bewick et al., 2010). Thus, we expect that *X. tropicalis* utilizes a different mechanism for sex determination than *X. laevis* does, though they could use an XX/XY or WZ/ZZ mode of sex determination. I investigated the mode of sex determination in *X. tropicalis* using a genetic approach, and undertook a genomic approach to finding the sex locus in this species.

II. Genetic approach

This part of the project was conducted in collaboration with Isabelle Philipp, with assistance from Honors undergraduate student Philip Auyang.

To investigate the mode of sex determination in *X. tropicalis*, we undertook a genetic approach that relied on the sex reversal of male frogs via estrogen treatment, based on the classic experiment that proved *X. laevis* uses a WZ/ZZ mode of sex determination (Figure 3.1A)(CHANG and WITSCHI, 1956). In the case of a WZ/ZZ mode of sex determination,

100% of the offspring of sex-reversed genetic males (ZZ females) and wildtype males (ZZ males) will be male. If on the other hand they use an XX/XY system, 66-75% of the offspring of sex-reversed genetic males (XY females) and wildtype males (XY males) will be male, depending on whether YY individuals are viable. Non-sex reversed genetic females will have 50% male and 50% female offspring when mated to wildtype males in either case. Thus, the ratios of male to female offspring from sex-reversed males should reveal the mode of sex determination.

Isabelle followed a protocol from the Zimmerman lab in the UK to estrogen-treat *X. tropicalis* frogs, upon reaching sexual maturity they were mated to wildtype males, and the sexes of the offspring were recorded. The male or female gonads of *Xenopus* froglets fixed in Bouin's solution can be identified by visual analysis soon after metamorphosis (Hayes et al., 2002). To make it easier to confidently distinguish the gonads, however, we typically waited at least a few weeks post-metamorphosis before dissecting them. Images of developing *X. tropicalis* gonads are shown in Figure 3.1B.

Table 3.1 shows the numbers and percentages of each sex for the offspring from each mother, for mothers that had 20 or more offspring. Although the numbers of offspring may not be high enough to achieve statistical significance, importantly, no mothers had 100% male offspring, as one would expect from sex-reversed genetic male mothers in a WZ/ZZ system. Furthermore, the offspring ratios fall into two classes: those with 50 +/– 9% males, and those with 60% or greater males. The way we interpret these results is that the first class of mothers represents genetic females and the second class of mothers represents sex-reversed genetic males. These data are inconsistent with a WZ/ZZ mode of sex determination, and are consistent with an XX/XY mode of sex determination.

Olmstead et al. reported having found sex-linked amplified fragment length polymorphisms (AFLPs) for *X. tropicalis*, but were not able to repeat their findings for all of their frogs (Olmstead et al., 2010). Isabelle and Philip were also unable to repeat the Olmstead et al. results, using the PCR primers given in the paper, or using primers Isabelle designed for simple sequence repeats (SSRs) on *X. tropicalis* genome v.4 scaffolds the Olmstead paper reported as being sex-linked. These results suggest that sex determination may be highly dynamic within *X. tropicalis*, and in fact one frog species within *Rana* has different populations that utilize different modes of sex determination (Hayes, 1998).

A

	WZ/ZZ mode (females heterogametic)		XX/XY mode (males heterogametic)	
estrogen -treated	WZ female	sex-reversed ZZ	XX female	sex-reversed XY
wildtype	x ZZ male	x ZZ male	x XY male	x XY male
offspring	50% WZ females 50% ZZ males	100% ZZ males	50% XX females 50% XY males	25% XX females 50% XY males 25% YY males OR 33% XX females 66% XY males

B

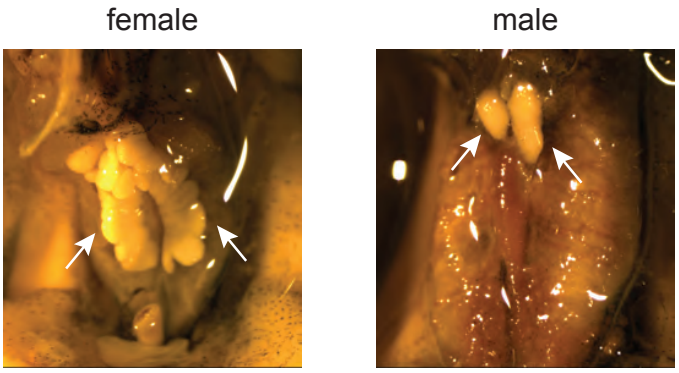


Figure 3.1. A genetic approach to finding the mode of sex determination in *X. tropicalis*. (A) Frogs are treated with estrogen during development, from tadpole stages through metamorphosis. Estrogen-treated genetic females (WZ or XX, depending on the mode of sex determination) will develop normally as females, and give the expected 50/50 ratio of male to female offspring when mated to wildtype males. Genetic males (ZZ or XY), on the other hand, will be sex-reversed by estrogen and develop as females. These sex-reversed genetic males will give different ratios of male and female offspring, depending on the mode of sex determination. There are two possibilities for the sex ratio from XY females because we don't know whether YY individuals would survive. (B) Frogs can be sexed by dissection a few weeks after metamorphosis. Ovaries or testes are marked with arrows.

Frog ID	Total offspring examined	# males	# females	percent male	percent female
C810-1 5R	99	69	28	70%	28%
C795-1/2 1R3R	73	51	20	70%	27%
C795-1/2 6G	45	27	15	60%	33%
C795-3/4 3Y	44	29	13	66%	30%
C810-2 3G	39	28	13	67%	33%
C810-4 2O5O	21	14	7	67%	33%
C810-1 5R6R	124	57	57	50%	50%
C810-1 3R5G	100	51	47	51%	47%
C795-1/2 2R	21	9	12	41%	55%
C810-4 2Y	20	10	10	50%	50%

Table 3.1. Numbers of sexed offspring for estrogen-treated frogs.

III. Genomic approach

This part of the project is in collaboration with the Rokhsar lab.

Using our GBS data from phenotyped F2 frogs (Chapter Two), Therese Mitros found that a region of *X. tropicalis* chromosome 7 correlates with sex (Figure 3.2). Genes that control sexual development tend to contain DM DNA-binding domains, for example, *Drosophila doublesex* and *C. elegans mab-3*, for which the DM domain is named; medaka *DMY*/*Dmrt1bY*, chicken *DMRT1*, and *X. laevis DMRT1* and *DM-W* all contain DM domains (Yoshimoto et al. 2008). Therefore, we decided to look for DM domain-containing genes in the correlated region, as candidates for the sex determining gene. Therese generated a list of protein domains coded for in the first 40 Mb of scaffold 7 using the PFAM algorithm, however, there are no DM domains encoded in this region.

Based on the reasoning that at the sex-determining locus, one sex is homozygous and the other heterozygous, we decided to take another approach to look for possible differences in read counts from males and females aligning at the sex locus. Using the Illumina Truseq kit, I made two sex-specific shotgun libraries from F13 Nigerian strain frogs, one from 12 pooled male DNAs and one from 8 pooled female DNAs (Table 2.1). Being 13 generations inbred, we assume that they are isogenetic, and therefore the two libraries should only differ at the sex locus. The frog whose DNA was used for the reference genome was a female. If *X. tropicalis* uses a WZ/ZZ mode of sex determination, we expect either the W or Z version of the sex locus to be in the assembly. In the former case, at the sex locus, reads from the female library would align but no male reads would align there. In the latter case, twice as many male than female reads would align at the sex locus. If *X. tropicalis* uses an XX/XY system, on the other hand, then the X version of the sex locus would be in the assembly, and we would expect twice as many female as male reads to align there. The data from these sex-specific libraries are currently being analyzed.

Recently, Therese re-analyzed the sex correlation using her new SNP map generated from our GBS data (see Chapter Two), and she narrowed down the region of correlation to marker super_547:1, at 8.96 centimorgans on chromosome 7, which corresponds to approximately 6–7 Mb on scaffold 7 of genome assembly version 7. At this marker, males either have the same genotype as the Nigerian male P0, or are heterozygous; females either have the same genotype as the ICB female P0 or are heterozygous. The fact that either sex can be heterozygous at this sex-correlated locus suggests that the genetics of sex determination in *X. tropicalis* is more complicated than we thought: it may be multigenic, may result from alleles with incomplete dominance over one another, and/or may incorporate some stochasticity. We are hopeful, however, that the data from the sex-specific libraries will help us parse these possibilities. In addition, we plan to investigating this putative sex locus more closely via PCR amplification and sequencing, and in more individuals. Doing so within each strain of *X. tropicalis* and comparing them to one another will give us insight into whether the different strains may be specifying sex in different ways.

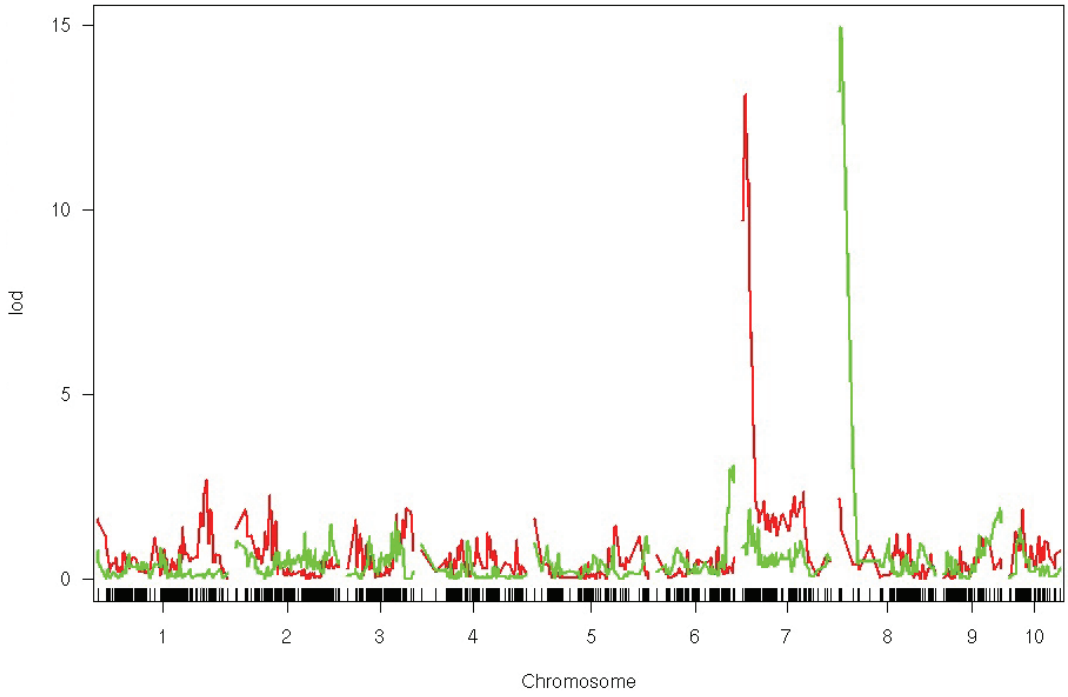


Figure 3.2. Sex (red line) is correlated with a region on *X. tropicalis* chromosome 7. Computation and figure by T. Mitros.

Chapter Four: Mapping and identification of the *curly* mutation in *X. tropicalis*

I. Background

Over a decade ago, a program was begun to develop *Xenopus tropicalis* as an amphibian genetic model organism (Amaya et al., 1998). *X. tropicalis* possesses almost all of the experimental advantages of the longstanding developmental, cell biological, and biochemical model organism *X. laevis* (Beck and Slack, 2001) but unlike the pseudotetraploid *X. laevis*, *X. tropicalis* is diploid (Hellsten et al., 2007). A number of background mutations emerged during inbreeding of *X. tropicalis* frogs in the Harland lab, including *curly* (Grammer et al., 2005). Gynogenesis, a method for obtaining diploid offspring with only maternal genetic material, allowed for the obtaining of mutant offspring in one fewer generation than would be required from conventional matings, and this technique was combined with genetic mapping using simple sequence length polymorphisms (SSLPs) to find linkage between mutations and chromosomes (Khokha et al., 2009).

Briefly, SSLPs are simple sequence repeats (SSRs, also known as microsatellites) that are of different length in different strains of a given species—in *X. tropicalis* we use the inbred Nigerian and Ivory Coast B (ICB) strains. Genetic mapping using SSLPs exploits the fact that DNA fragments PCR-amplified using primers flanking an SSLP can be genotyped by resolution via polyacrylamide gel electrophoresis. The more tightly linked two markers are to one another, the less likely it is that a recombination will occur between them and thus the less likely it is that they will have different genotypes in the same individual. Similarly, the more closely linked a mutation is to a given marker, the greater proportion of mutant individuals should be homozygous for the marker characteristic of the strain in which the mutation arose (in this case the Nigerian strain).

Xenopus SSRs are highly AT-rich, and CG repeats are rare (Xu et al., 2008). On average, each megabase of *X. tropicalis* genome sequence contains 161 dinucleotide, 27 tetranucleotide, and 17 trinucleotide microsatellites with a minimum of 5 repeats (Xu et al., 2008). Although it is not known how these microsatellites arise, Xu et al. showed that dinucleotide repeats are the most common, and longer repeats are more likely to be polymorphic.

The Sater lab at the University of Houston characterized a panel of over 2800 SSLP markers derived from di-, tri-, or tetranucleotide SSRs identified computationally. Markers shown to be polymorphic were amplified from 190 F2s using fluorescent PCR primers and resolved via capillary electrophoresis. The resulting genotype data was used to construct a genetic map of *X. tropicalis* (referred to herein as "Sater markers") (Wells et al., 2011). Due to strain differences, only a subset of these markers has proved usable for our mapcrosses and our experimental conditions. Nevertheless, having this list to work from has been invaluable. For example, as explained below, using Sater markers plus only two custom markers we have been able to map *curly* to a window of 1.9 Mb.

II. Genetic mapping of *curly*

As noted above, *curly* is a recessive background mutation discovered in the Nigerian lineage of *X. tropicalis*. It follows a Mendelian inheritance pattern and is embryonic lethal. The *curly* phenotype is characterized by a ventral edema and a kinked or curled tail (Grammer et al., 2005). The phenotype is first evident as a dorsal curvature and mild ventral edema around the heart starting in the mid-30's tadpole stages. As development proceeds, the edema increases and the tail bends to an increasing degree, sometimes curling (Figure 4.1A). The gut does not develop properly, and *curly* tadpoles do not live past the late 40's feeding tadpole stages.

Toral Trivedi, as part of her undergraduate honors thesis in the Harland lab, mapped the *curly* mutation to the q arm of chromosome 4 (Khokha et al., 2009; Trivedi, 2009), which corresponds to scaffold 4 of *X. tropicalis* genome assembly version 7. She and Dipankan Bhattacharya obtained a mapping population of approximately 1800 *curly* mutant tadpoles from a single natural mating. Their DNAs as well as those of 48 of their wildtype siblings were isolated via alkaline lysis (Figure 4.1B) (Bhattacharya, 2010). The relatively easy procurement of such a large number of mutant embryos speaks to the strength of the *X. tropicalis* system. The *X. tropicalis* mutant *dicky ticker*, for example, was successfully mapped to a 230 kb window using only 562 mutant embryos (Geach and Zimmerman, 2010), so we expected that 1800 should be sufficient for fine mapping of *curly*.

Honors undergraduate student Raha Sadjadi and I proceeded with the intermediate mapping of *curly* on chromosome 4 using Sater markers (Wells et al., 2011) (Figure 4.2A, Table 4.1). As described in Wells et al., the genetic map of those markers is often inaccurate in the relative placements of markers near one another, with markers on adjacent but distinct v.4 scaffolds being intermixed for reasons that are unknown. Taking this into account, I decided not to use the genetic map per se but rather to use a marker order derived from *X. tropicalis* genome assembly version 7.1 (assembled by Jeremy Schmutz and Jerry Jenkins; currently accessible at xenbase.org), via BLAST alignment of the primers to the assembly (A. Session and J. Jenkins, personal communications). It was possible for me to use such a marker order for mapping because the scaffolds in version 7 are at a chromosome scale. As shown in Table 4.1, our mapping data show that only two of the markers in the region of our mutation were placed incorrectly relative to one another by using the order from the genome assembly.

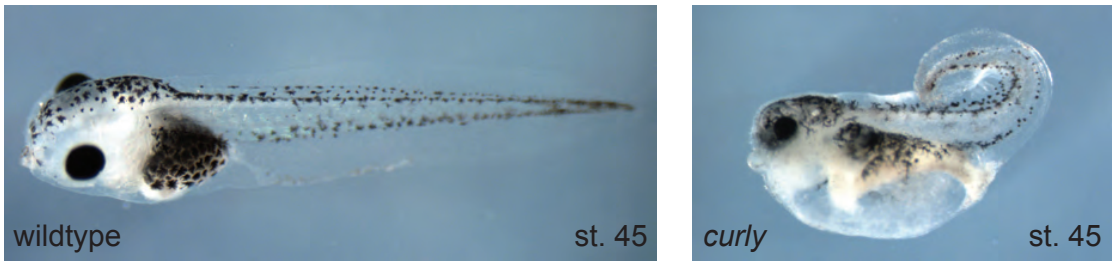
Since the *curly* mutation arose in the Nigerian strain, we expect *curly* mutant tadpoles to be homozygous Nigerian at the mutant locus. This means that as I approached the mutant locus, I would sample fewer and fewer non-Nigerian genotypes. Indeed, as we proceeded along scaffold 4, we found fewer and fewer heterozygous tadpoles (Figure 4.2A, Table 4.1). As we recovered fewer and fewer non-Nigerian genotypes from mutant DNAs, we relied on the wildtype DNAs as positive controls for the resolution of genotypes from alkaline lysis-isolated DNAs. Initially, to find the general region where *curly* is located, I had genotyped a series of markers using only 48–96 DNAs from the "A" set, moving along the chromosome away from the centromere (toward the right in Table 4.1) and none of these markers appeared to be distal to the mutation. Once Raha joined me, our strategy was that she would genotype more individuals for our current most proximal marker (at various points, 019A05, 019B11,

031D02) to recover more useful recombinants, as I tested putative distal markers to see if they would be usable for our mapcross. Later, the useful recombinants from proximal markers (e.g. B94 and B80) were tested at other proximal markers (e.g. 044E10 and 008H02) and putative distal markers we later identified (e.g. Sc4_107174736) to confirm the relative placements of the markers, and that the recombinant individuals genotyped as expected.

A few individuals had heterozygous or ICB genotypes spanning our entire mapping region, from 044E10 to 107G02 (Figure 4.2A, data not shown), but I excluded them with the interpretation that they must have been mistakenly identified as mutants during the DNA isolation, since all other mapping data consistently pointed to the mutation lying in that region. Presumably, these individuals were mistaken for mutants because they had a nonspecific developmental defect that caused them to resemble *curly* embryos; for example, edema is relatively common in developing *X. tropicalis* tadpoles. Once these missorted individuals were excluded, it became clear that 031D02 is in fact distal to the *curly* mutation, because the six heterozygous tadpoles (out of 1034 genotyped) we found at that marker are homozygous Nigerian on the proximal side, indicating that they must have undergone a recombination distal to the *curly* locus but proximal to 031D02 (Table 4.1). These recombinants may be useful for finer mapping in the future.

Once we determined that 031D02 was in fact a distal marker, we tested markers between 019B11, our current proximal SSLP marker from the Sater map, and 031D02. Although none of the Sater markers that lie between 019B11 and 031D02 were usable for genotyping our mapping population, one new primer set, s350-936874 (designed by I. Philipp) was usable and allowed us to narrow our mapping window to a 1.9 Mb window on scaffold 4 (Figure 4.2A, Table 4.1). Out of the 11 heterozygote individuals we had found at 019B11, only one of those remains heterozygous at this proximal boundary of our mapping window. Our most relevant mapping results are shown in Table 4.1.

A



B

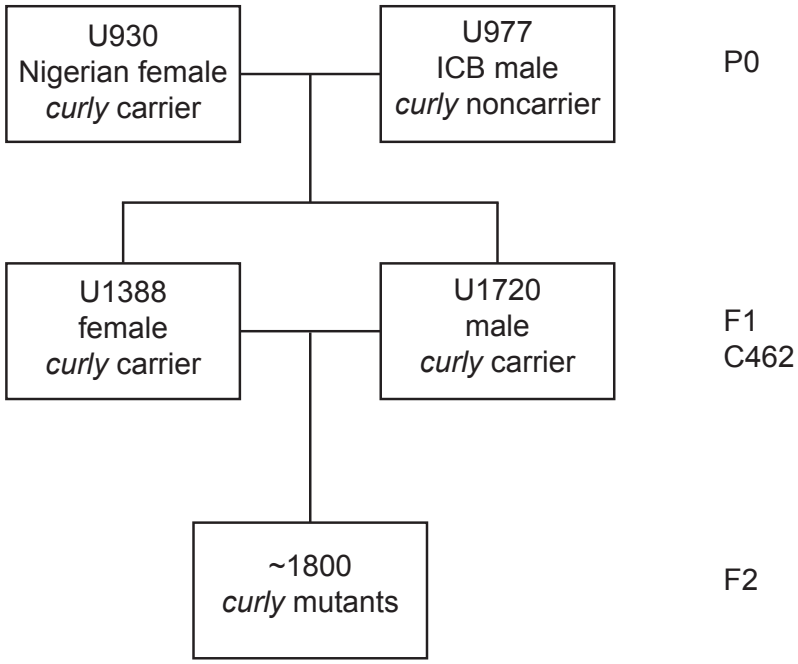


Figure 4.1. *curly* phenotype, and frogs used for mapping. (A) The *curly* phenotype is characterized by a ventral edema and curled tail. (B) The *curly* mutation arose in the Nigerian strain. The lineage of the approximately 1800 *curly* tadpoles generated from a mapcross is shown.

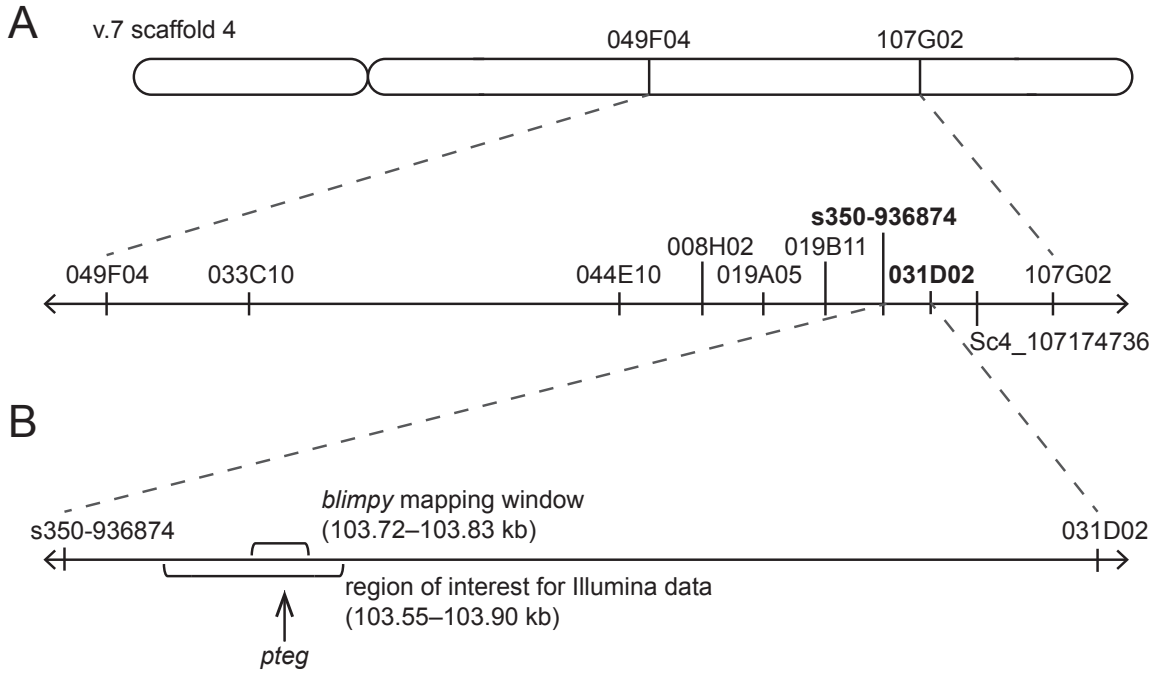


Figure 4.2. Mapping landscape for *curly*. (A) Relative positions of selected markers used for mapping *curly*. Markers s350-936874 and 031D02, indicated in bold, define the *curly* mapping window. (B) Relative positions of the *blimpy* mapping window, the region used for Illumina sequence analysis, and the *pteg* gene within the *curly* mapping window.

Marker:	049F04	033C10	044E10	008H02	019A05	019B11	**s350-936874**	**031D02**	Sc4_107174736	107G02
v.7 sc 4 coordinate:	72.3 Mb	78.0 Mb	92.8 Mb	96.1 Mb	103.0 Mb	101.0 Mb	103.3 Mb	105.3 Mb	107.1 Mb	110.2 Mb
Total F2s genotyped:	92	45	100	122	572	570	318	1034	284	223
Sample										
D27					N	F, F	N	H		
C18			N	N	N	N	N	H	H	H
F51					N	N		H		
F52					N	N		H		
F75					N	F		H		
G86			N	N			N	H	H	H
F62			H	H	H	H	H	N, N		
A16	H	H	N	N	N	N		N	H	H
A18	H	H	N	N	N	N		N	H	H
A23	H	H	N	N	N	N		N	H	H
A64	N			H	H	H	N	N, N	N	N
A74	H				N	N		N	H, H	H
A80	N		H	H	H	H	N	N, N	N	N
A88	N			H	H	H	N	N, N	N	N
B1			H	H	H	H	N	N, N	N	N
B28			H	H	H	H	N	N, N	N	N
B65			H	H	H	H	N	N, N	N	N
B45					N	N		N	H	H
B70					N	N		N	H	H
B84					N	N		N	H	H
B94			H	H	H	H	N	N, N	N	
B80			H	H	H	H	N	N, N	N	N
B95					N	N		N	H	H
C5			H	H	H	H	N	N, N	N	
C20			N	N	N	N	N	N	H	H
C83			H	H	H	H	N	N, N		
C62					N	N	N	N	H	H
A43	H	H	N	N	N	N		N	N, N	H
A58	H			N	N	N		N	N, N	H
A90	H				N	N		N	N, N	H
B17					N	N		N	N	H
A77	N		H	H	H, H	N, N	N	N, N	N	N
B72			H	H	H, H	N, N	N	N, N	N	N
A2	N	N	H	H	N	N		N	N	N
A66	N			H	N	N		N	N	N
D49			H	H	H	N, N		N, N		
D79			H	H	H	N	N	N, N		
E58					H, H	N, N		N, N		
A4	N	N	H	N	N	N		N	N, N	N
A10	N	N	H	N	N	N		N	N	N
A6	H	H	N	N	N	N		N	N	N
A8	H	H	N	N	N	N		N	N	N
A17	H	H	N	N	N	N		N	N	N
A24	H	H	N	N	N	N		N	F, N	N
A28	N	H	N	N	N	N		N	N	N
A30	H	H	N	N	N	N		N	N	N
A33	H	H	N	N	N	N		N	N	N
A94	N	H			N	N		N	N	N
A83	H				N	N		N	N	N
A86	H				N	N			N	N

Table 4.1. SSLP genotyping results for useful individuals at selected markers. "N" indicates a homozygous Nigerian genotype, "H" indicates a heterozygous Nigerian/ICB genotype, "F" indicates a failed PCR, and a blank box indicates that the individual was not genotyped at that marker. Where two genotypes are given, they represent the results of two separate reactions. Our data support the relative positions of 019A05 and 019B11 as shown, as opposed to that based on their v.7 coordinates. The markers flanking our mapping window are in bold. All markers are from (Wells et al., 2011), except s350_936874 and Sc4_107174736.

III. *curly* and *blimpy*

I noticed that the *curly* mapping window encompassed the mapping window of *blimpy*, a mutation Isabelle Philipp was mapping and which had been thought to be an induced mutation (the relative positions of the two mapping windows is shown in Figure 4.2B). The *blimpy* carriers are descended from frogs exposed to gamma irradiation, but who were from the Nigerian strain, and since their mutant offspring develop a ventral edema and kinked tail reminiscent of the *curly* phenotype, it seemed possible that *blimpy* frogs in fact carried the *curly* mutation.

To test whether *curly* and *blimpy* frogs could be carrying the same mutation, Isabelle and I performed a complementation test between a male *curly* carrier and a female *blimpy* carrier. Isabelle counted that 26% of the offspring from this cross displayed a phenotype very similar to the *curly* phenotype (Figure 4.3A), showing that the mutations causing the *curly* and *blimpy* phenotypes fail to complement. Although it is formally possible that the two lines could carry different mutations that affect the same gene, this result is consistent with both phenotypes being caused by the same mutation.

Isabelle had previously shown that the gene *pteg* (also known as *pdzk1ip1*), one of three in her mapping window, is misspliced in *blimpy* mutants inasmuch as exon 3 is missing from *blimpy pteg* transcripts although it is not missing from the genomic DNA. To test whether *pteg* is misspliced in *curly* embryos, Raha performed RT-PCR for *pteg* on a pool of 10 *curly* embryos and a pool of 10 of their phenotypically wildtype siblings, using Isabelle's primers. In the wildtype lane, she saw an upper band at about the expected size of 342 bp, as well as a smaller band. In the *curly* lane, only the smaller band was amplified (Figure 4.3B). I gel extracted and sequenced the amplified bands and confirmed that the smaller band seen in both lanes represents *pteg* transcripts lacking exon 3, and the larger band corresponds to the wildtype form of *pteg*. As expected, both versions of the transcript were amplified from the pool of phenotypically wildtype embryos because this pool would comprise *curly* noncarriers as well as heterozygotes.

The loss of exon 3 from *pteg* transcripts, as in *curly* mutant embryos, does not result in a frameshift; rather it results in a loss of 28 amino acids from a region with no predicted protein domain or motif (Figure 4.4; SMART: http://smart.embl-heidelberg.de/; Lee et al. 2010). Isabelle has shown via *in situ* hybridization, however, that levels of *pteg* mRNA are reduced in the kidney (the only place where *pteg* expression is robustly detected) in *curly* mutants, so the question of whether the *curly* form of the protein would be functional may be moot.

Based on the derivation of the *blimpy* line from the Nigerian strain, the coincidence of the *curly* and *blimpy* mapping windows, the failure of *curly* and *blimpy* to complement one another, and the fact that *pteg* is misspliced in both lines, the most parsimonious view is that the *blimpy* and *curly* phenotypes are in fact caused by the same mutation. Since the *blimpy* carriers are descended from frogs exposed to a mutagen, it remains formally possible that they carry an induced mutation in addition to the one that affects *pteg*. It is unlikely, however, that such a mutation contributes to the *blimpy* phenotype, because it would have to

remain linked to the *pteg* locus through several meioses, and the likelihood of an induced mutation occurring immediately adjacent to a preexisting mutation must be vanishingly small.

I will hereafter refer to the *curly*/*blimpy* mutation and phenotype as *curly*, except when pointing out a difference between the two lines.

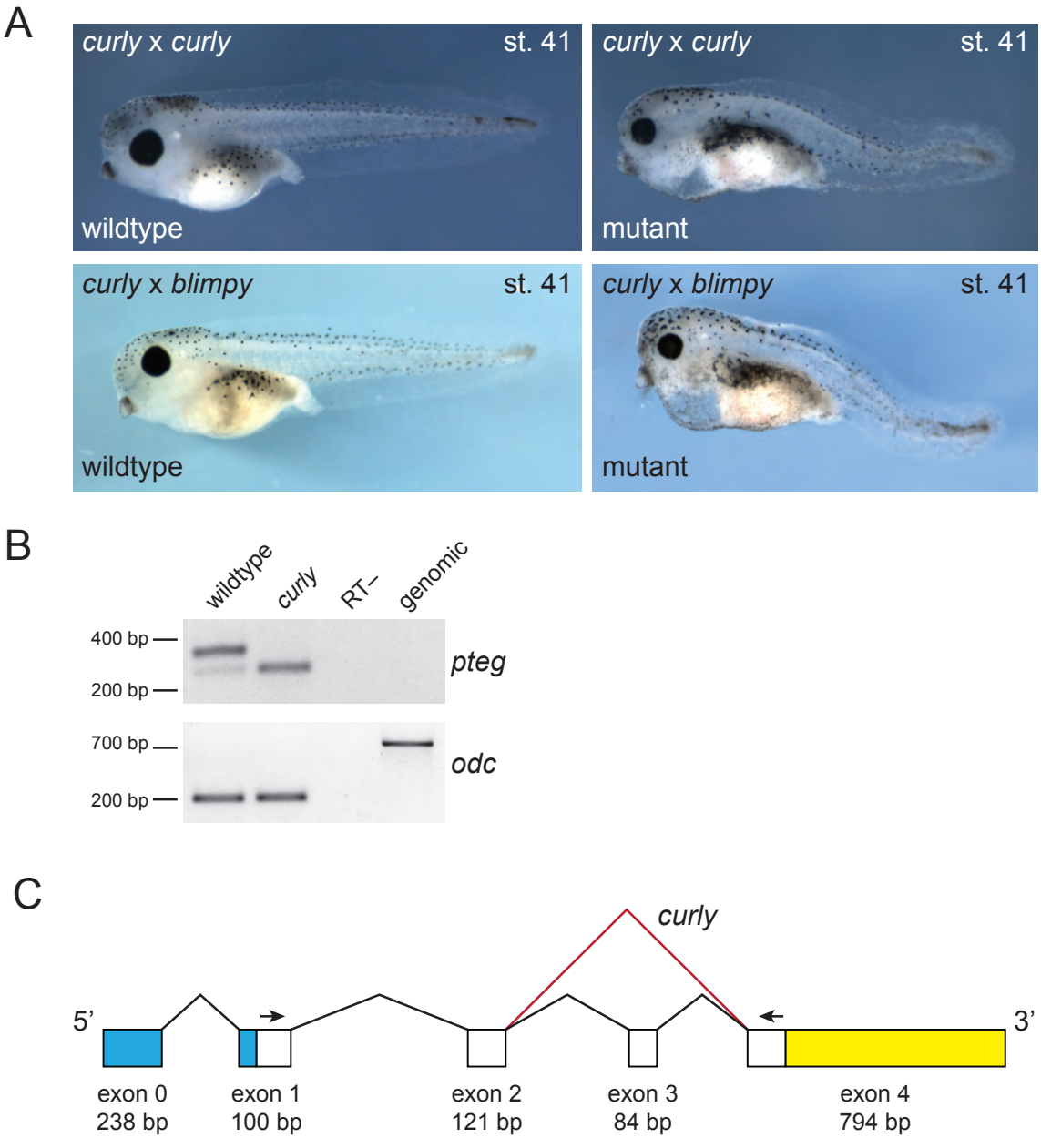


Figure 4.3. *curly* and *blimpy* fail to complement, and *pteg* transcripts are misspliced in *curly* mutants. (A) When *curly* and *blimpy* mutant carriers are crossed, approximately 1/4 of the offspring display the ventral edema and kinked tail characteristic of the *curly* phenotype. (B) RT-PCR for *pteg* and the loading control *odc* was performed by R. Sadjadi, using cDNA from pools of 10 curly embryos or 10 phenotypically wildtype siblings (the latter comprising homozygous wildtype as well as heterozygous genotypes). (C) Architecture of the *pteg* gene. The 5'UTR is in blue and the 3'UTR is in yellow. Arrows denote the locations of the PCR primers used for *pteg* in B. In *curly* embryos, as in *blimpy* embryos, *pteg* transcripts lack exon 3.

A

```
atgttatccctgcaacatttgctccttattctttttccttggggcaggtctctgcacagcacgttcat
 M  L  S  L  Q  H  L  L  L  I  L  F  S  L  G  Q  V  S  A  Q  H  V  H

aataatgttggcagaagatttcctcagtggttaacagggcttatcgctatgacagtcttcctctttctc
 N  N  V  G  R  R  F  P  Q  W  L  T  G  L  I  A  M  T  V  F  L  F  L

gttcttgtggtatatgtagccaagatgctctggaaaaagagatcacagcaaggcacaaacatgaaagac
 V  L  V  V  Y  V  A  K  M  L  W  K  K  R  S  Q  Q  G  T  N  M  K  D

tttgaagaagtggtcgccaatggcaccggtggatgttatgagacaagaatagaaaatatctggtccggt
 F  E  E  V  V  A  N  G  T  G  G  C  Y  E  T  R  I  E  N  I  W  S  G

gaaaacattcatgcatatgaaaatcccattgaagtcaatgacaatgtccgtactacagctatg
 E  N  I  H  A  Y  E  N  P  I  E  V  N  D  N  V  R  T  T  A  M
```

B

```
wildtype   1 MLSLQHLLLLILFSLGQVSAQHVHNNVGRRFPQWLTGLIAMTVFLFLV
curly      1 MLSLQHLLLLILFSLGQVSAQHVHNNVGRRFPQWLTGLIAMTVFLFLV

wildtype  48 LVVYVAKMLWKKRSQQGTNMKDFEEVVANGTGGCYETRIENIWSGEN
curly     48 LVVYVAKMLWKKRSQ----------------------------SGEN

wildtype  95 IHAYENPIEVNDNVRTTAM
curly     95 IHAYENPIEVNDNVRTTAM
```

C

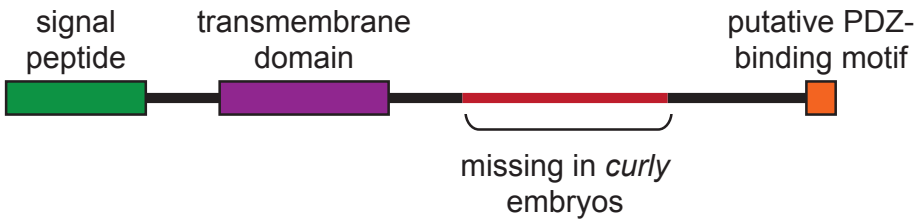


Figure 4.4. *pteg* sequence and domain structure. (A) Coding sequence of *pteg*, and corresponding amino acid sequence. Exon 3, missing from *curly* transcripts, is in red. (B) Protein sequence alignment for wildtype and *curly* forms of *pteg*. The wildtype form consists of 113 amino acids, and the *curly* form has 85 amino acids. (C) Domain structure of *pteg*, based on SMART annotation and (Lee et al., 2010).

IV. Illumina sequencing

Isabelle had previously performed Sanger sequencing on almost all of the *pteg* locus, and was unable to locate any differences between mutant DNA and the Nigerian reference genome, so we decided to use the Illumina platform to sequence the genome of a pool of mutant tadpoles from the *blimpy* line. (At the time our mapping data indicated that the *blimpy* and *curly* mapping windows did not overlap, and so we thought that the *blimpy* frogs carried *curly* in addition to another mutation and that we'd capture both this way). Isabelle sorted and extracted the DNA from a pool of 100 *blimpy* tadpoles, and we constructed a library using the Illumina TruSeq kit per the kit instructions. This library was sequenced as 100 bp paired-end reads on a HiSeq 2000 by the Vincent Coates Sequencing Center at UC Berkeley.

Isabelle aligned the sequencing reads to *X. tropicalis* genome assembly version 7 using computational methods. Notably, she allowed up to 15 mismatches per read, with the idea that most reads containing SNPs and/or indels would still align. The alignment of the reads to the reference resulted in what is known as an mpileup file, which, for each position in the reference, gives the base in the reference and the basecall for each sequence read aligning there. It also gives a quality score for each read, which takes into account the quality score of the sequencing as well as the quality of the alignment to the genome (Baq calculation; T. Mitros, personal communication).

Our next step was to examine the mpileup file described above. Rather than looking globally for potential mutant loci, we decided to focus on the region most likely to contain the *curly* mutation, based on our mapping data. We thus defined a "region of interest" wider than the narrowest *blimpy* mapping window, but bordered by the markers in which Isa had the most confidence. This 350 kb region is well within the *curly* mapping window (Figure 4.2B).

I evaluated the mpileup file in our region of interest by visual inspection. Since the DNA sequenced was from pooled *blimpy*/*curly* mutants, we expected that at the site of the lesion, they should be homozygous and different from the reference. Since neither the frog used for the reference genome nor the blimpy carriers are very inbred, there will be places that differ between their genomes that are not associated with the *curly* phenotype and thus we expected a number of false-positives from this analysis. We didn't, however, expect any false negatives (except in the case of sequencing error), because if a given lesion is causative of the *curly* phenotype then we didn't expect this locus to be shared with the reference. In my evaluation of potential loci of interest, I used my judgment to gauge whether a given locus was homozygous and different from the reference, taking the amount of coverage and the Baq-calculated quality scores into account. Via manual curation, I identified 523 homozygous SNPs in 349,999 bp, which is equivalent to 1 SNP per 669 bases (Table 4.2). I also identified 86 indels, which are listed in Table 4.3.

Upon the revelation that the frog used for the genome assembly left no descendants (M. Khokha, personal communication), we became concerned that perhaps she had carried *curly*, which would certainly confound the analysis of our *curly* Illumina data. Also, because she was not very inbred (approximately seven generations) (Hellsten et al., 2010), simple

polymorphism between the *blimpy* line and the reference probably contributed to the high number of differences we found in our Illumina data. Thus, we decided to take advantage of the Nigerian F13 shotgun data I generated (see Chapter 2): any locus where the F13 data also differ from the reference can't be lesions that cause *curly*, so we will subtract those places from our list of differences. Isabelle is currently performing this analysis.

Since the only known differences in *cis* that can affect the splicing of a transcript are within or very near (conservatively, within 2 kb of) the affected gene (D. Rio, personal communication) (Kornblihtt, 2005), we decided to first evaluate the differences within or near the *pteg* gene. Those within 10 kb of the *pteg* transcript are shown in Figure 4.5. Our first filter for evaluation of whether a given difference could be causative of the *curly* phenotype is based on its inheritance pattern: if a given SNP or indel is causative, then we expect all mutants (from both the *curly* and *blimpy* lines) to be homozygous for that SNP or indel, all carriers to be heterozygotes, and related noncarriers as well as inbred frogs not to have the lesion. Efforts by Raha to characterize the putative lesions near the *pteg* gene via Sanger sequencing according to these criteria are ongoing.

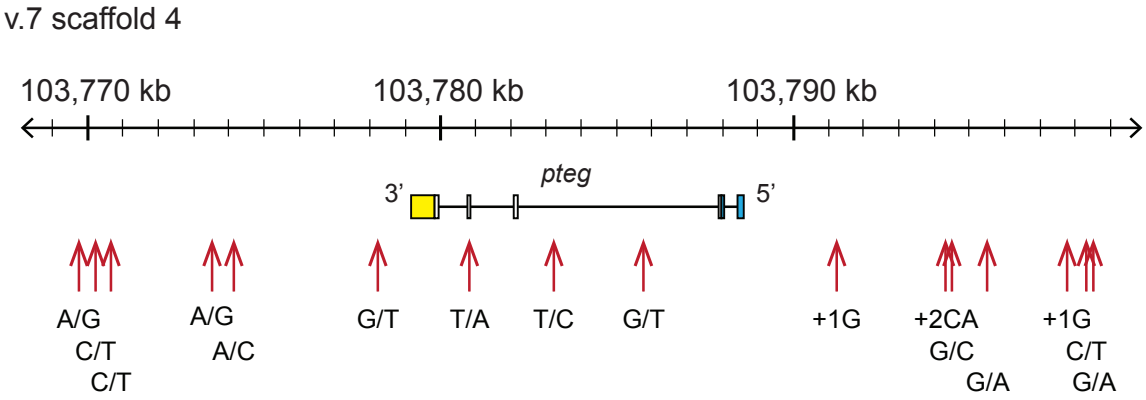


Figure 4.5. Places within 10 kb of the *pteg* gene where Illumina-sequenced mutant frogs differ from the reference. SNPs are denoted as X/Y, where X is the base in the reference and Y is the sequenced base. Insertions are denoted with a +, followed by the number of bases inserted and the number of bases. No deletions were identified within 10 kb of the *pteg* transcript. The 5' and 3' UTRS of the *pteg* transcript, labeled in blue and yellow, respectively, are based on data from RACE performed by I. Philipp.

Table 4.2. Putative SNPs between *curly* embryos and the reference, based on manual curation of Illumina sequence for our region of interest.

v.7 scaffold 4 coordinate	Base in reference	Sequenced base
103554094	G	A
103557234	A	G
103557236	A	G
103557316	A	C
103557382	A	T
103579428	A	T
103565310	C	G
103575485	T	C
103579459	T	A
103579460	G	A
103579865	C	A
103582739	G	A
103582799	A	G
103583065	C	G
103590353	T	C
103590988	G	A
103592477	T	C
103593395	A	G
103598337	T	C
103613367	G	A
103634521	A	C
103634621	T	A
103635547	C	T
103635565	A	G
103635767	A	G
103635768	A	T
103635770	A	T
103635774	T	C
103635808	A	C
103635825	A	G
103639691	T	A
103639693	C	T
103639852	T	A
103652788	T	C
103652789	T	G
103652790	T	C
103652791	T	A
103653630	T	G
103656519	G	A
103658351	T	G
103664182	T	G
103664183	A	T
103664224	T	C
103664225	C	A
103664274	C	G
103664761	A	T
103668581	G	A
103668656	A	G
103668781	T	C
103668783	A	G
103668874	A	C
103668886	G	A
103668907	C	A
103668963	C	T
103668979	G	A
103669016	T	C
103669017	C	T
103669071	A	G
103669073	G	A
103669097	C	G
103669099	C	G
103669105	T	C
103669339	T	C
103669786	A	C
103669848	T	C
103670285	T	C
103670338	G	A
103670341	G	A
103670461	A	G
103670540	C	G
103671193	T	A
103671196	C	G
103671335	C	T
103671338	G	C
103671828	A	C
103671844	G	C
103672217	G	A
103673017	C	T
103673079	T	G
103673743	T	G
103674771	C	A
103674772	C	A
103674774	G	A
103675227	C	T
103675291	A	G
103675318	A	T
103675871	T	C
103675933	G	T
103678844	T	C
103679063	C	T
103679087	G	A
103679176	A	T
103679253	C	G
103679482	T	A
103680517	T	A
103684018	A	G
103686640	A	G
103686863	C	T
103686886	G	A
103686888	G	A
103687274	A	G
103687315	T	G

Table 4.2, continued.

103688744	T	G
103689135	C	A
103689890	G	A
103690154	G	T
103690226	T	G
103690690	G	C
103690691	G	C
103690701	T	C
103690950	C	T
103690960	G	A
103690962	A	T
103691857	A	T
103691895	G	T
103693385	T	C
103696698	G	T
103696760	G	A
103697595	T	C
103697666	G	A
103697685	T	C
103697848	A	C
103698034	T	A
103698038	G	A
103698245	T	C
103698264	G	A
103698329	C	A
103698363	T	C
103698449	A	G
103698605	C	T
103699186	T	G
103699672	G	C
103699676	T	C
103699690	C	T
103699693	C	T
103699750	G	T
103699789	A	G
103700124	A	G
103700162	G	T
103700475	A	C
103700492	C	A
103700605	T	C
103700625	T	C
103701082	A	T
103701083	A	T
103701089	T	A
103704756	A	C
103704983	A	G
103707256	G	T
103707419	G	A
103712557	A	C
103712575	G	C
103712657	G	C
103713113	T	C
103713968	A	T
103715177	T	C
103715197	A	G
103715337	T	G
103715697	T	C
103716520	G	A

103718435	A	G
103718501	A	T
103719860	A	T
103721156	G	A
103722431	T	G
103722748	A	G
103724001	A	G
103724296	T	C
103724669	A	G
103724670	A	T
103724676	T	G
103724679	T	C
103724681	A	C
103724682	T	A
103724686	G	A
103724691	G	A
103724692	T	A
103724697	C	G
103724702	T	G
103724847	T	C
103725242	G	A
103725905	A	C
103726340	G	A
103726471	C	A
103726944	G	A
103726978	G	C
103728072	T	G
103728466	G	C
103728515	T	G
103728790	T	A
103729119	T	A
103729374	G	A
103729393	A	C
103729518	G	C
103732476	G	A
103732569	G	A
103732775	T	C
103733692	A	T
103733699	G	T
103733719	C	T
103734285	A	G
103735302	T	A
103735434	C	T
103735495	T	A
103735944	T	C
103737799	G	T
103737845	T	A
103737846	A	C
103737860	A	G
103737893	C	T
103737955	T	A
103737961	A	G
103738082	A	G
103738092	T	G
103738121	A	T
103738122	G	A
103739100	T	C
103739141	A	T

Table 4.2, continued.

103739156	A	C
103739158	C	A
103739162	C	T
103739168	A	G
103739207	G	C
103740128	C	T
103740142	C	T
103740149	G	A
103740162	C	T
103740177	A	C
103740180	G	T
103740233	A	T
103740264	C	T
103740283	G	C
103740287	C	T
103740301	A	G
103740341	C	T
103740370	T	C
103740411	T	A
103740413	T	C
103740417	G	A
103740426	T	C
103740427	T	G
103740464	C	T
103740486	C	G
103740529	A	T
103740544	T	C
103740603	G	C
103740607	C	T
103740611	G	C
103740633	G	A
103740640	G	A
103740645	C	T
103741075	T	A
103741084	G	A
103741103	A	G
103741115	G	A
103741152	G	C
103741185	T	C
103741189	A	T
103741191	G	A
103741194	A	T
103741196	T	A
103741229	A	G
103741260	G	C
103741282	A	G
103741322	A	C
103741340	T	A
103741481	G	A
103741498	G	T
103741509	A	C
103741524	A	G
103741559	C	T
103741565	C	A
103741579	G	T
103741600	C	T
103741637	A	G
103741672	C	T

103741703	T	A
103741710	A	G
103741723	A	G
103741724	T	G
103742066	A	G
103742071	G	T
103742075	T	C
103742103	C	G
103742471	G	T
103742721	C	A
103742831	T	C
103742846	C	T
103742848	A	C
103742878	T	C
103742896	T	A
103742907	T	G
103742931	C	T
103742942	C	T
103742952	A	G
103742987	A	T
103743456	A	G
103743469	A	G
103743483	G	T
103743484	C	T
103743498	T	C
103743499	G	T
103743502	T	C
103743506	A	G
103743507	C	T
103743532	T	C
103744499	A	G
103744511	A	C
103744514	A	T
103744555	T	C
103744605	C	A
103744626	T	C
103744653	G	A
103744657	G	T
103744661	G	C
103744717	A	C
103744727	G	A
103744810	G	A
103744829	A	C
103744853	A	G
103744903	C	A
103744911	T	C
103744949	G	A
103745012	C	A
103745053	T	C
103745104	C	T
103747699	A	T
103747878	A	G
103754870	G	T
103756832	G	A
103762750	C	T
103763066	C	T
103763077	C	A
103763368	G	A

Table 4.2, continued.

103764200	G	A
103765350	C	G
103765696	A	G
103766406	T	C
103766569	C	T
103766602	G	A
103766956	G	A
103769703	A	G
103770107	C	T
103770747	C	T
103773538	A	G
103774049	A	C
103778212	G	T
103780837	T	A
103783251	T	C
103785802	G	T
103794220	G	C
103795554	G	A
103797606	A	C
103798333	C	T
103798430	G	A
103800194	G	T
103800264	C	G
103800265	C	A
103800268	T	A
103800715	G	T
103801676	G	C
103803839	G	T
103805151	A	C
103812932	T	A
103812984	T	G
103813029	G	C
103813047	A	T
103814093	A	G
103814096	C	A
103814108	A	T
103814109	G	A
103814226	G	A
103814640	C	A
103814662	G	A
103814713	G	A
103815092	C	T
103815127	T	C
103816403	T	G
103817500	C	T
103820927	T	A
103824477	C	T
103824555	T	G
103824593	T	C
103824630	T	A
103827750	G	T
103828508	G	T
103828516	C	A
103828517	T	C
103828690	T	C
103828859	T	C
103829475	C	T
103830056	A	G

103830650	C	T
103831092	T	C
103831622	G	A
103831745	A	G
103831855	A	G
103834878	T	C
103834884	T	A
103835173	T	C
103835261	G	C
103835264	G	A
103835531	G	A
103835598	C	T
103835802	G	C
103835848	C	A
103835887	T	C
103835897	A	C
103836441	T	C
103836670	C	A
103836900	T	A
103837212	T	C
103837386	A	G
103837387	A	T
103837530	A	G
103837576	G	T
103837578	G	T
103837596	T	C
103837597	T	A
103837607	T	C
103837608	T	G
103837609	T	G
103837639	G	A
103837647	G	A
103837663	A	C
103837672	T	C
103837702	C	T
103837712	T	A
103837713	T	A
103837733	G	T
103837809	A	G
103837852	G	A
103837873	C	T
103838024	A	G
103838297	C	T
103841609	C	G
103842131	C	A
103842325	A	C
103842337	C	T
103842555	A	G
103842834	T	C
103842894	G	A
103842900	T	A
103842932	A	G
103843790	A	G
103843797	G	C
103843807	T	A
103843814	C	A
103843848	C	T
103843870	C	T

Table 4.2, continued.

103844018	A	T
103844044	A	T
103844056	G	A
103844127	A	T
103844148	A	G
103844317	G	A
103844384	C	G
103844592	G	A
103844863	C	G
103845259	T	C
103845303	C	T
103845428	C	G
103845593	C	A
103845594	A	G
103845819	G	A
103845996	G	T
103846116	T	A
103846521	T	C
103846627	A	G
103846721	A	C
103846886	C	G
103846923	T	A
103846929	G	T
103846945	A	G
103846946	T	G
103846950	A	T
103846995	C	A
103848503	A	G
103848869	A	G
103848906	G	C
103849084	A	G
103849212	T	C
103849246	A	G
103849498	A	G
103850001	T	G
103851264	C	G
103851440	C	G
103854322	C	G
103873274	A	G
103884880	A	T
103889371	C	T
103889641	T	C
103889660	C	T
103890360	A	G
103890377	A	G
103890429	A	T
103890806	A	C
103890811	T	C
103891642	A	C
103891680	T	A
103892211	T	C
103892608	T	A
103892611	T	G
103893009	C	A
103893036	G	A
103893041	T	G
103895058	A	T
103895635	A	G
103895667	T	C
103896412	C	T
103896491	T	G
103896505	T	A
103896515	A	T
103896976	G	A
103896979	C	G
103897170	T	C
103897567	C	T
103898015	G	T
103898821	T	A
103898824	A	T
103898826	A	T
103899270	C	A
103899390	T	A

Table 4.3. Putative indels between *curly* embryos and the reference, based on manual curation of Illumina sequence for our region of interest.

v.7 scaffold 4 flanking coordinates	Type	Sequence
103634554–103634556	deletion	G[-1A]T
103644107–103644114	deletion	A[-6AGCAAA]T
103671340–103671341	insertion	A[+2GG]C
103671356–103671357	insertion	G[+3AAA]A
103672773–103672774	insertion	C[+4TAGA]A
103673077–103673079	deletion	A[-1T]T
103674773–103674774	insertion	A[+1T]G
103690009–103690012	deletion	C[-2TG]A
103690472–103690473	insertion	T[+18ATGGGTGTTATGACACCA]T
103690681–103690684	deletion	T[-2TG]C
103690704–103690707	deletion	T[-2TG]C
103691734–103691742	deletion	T[-7CAGGTGC]C
103697982–103697983	insertion	C[+2AG]A
103703379–103703380	insertion	G[+3CCT]C
103704980–103704981	insertion	A[+1T]G
103705016–103705017	insertion	C[+1A]A
103713970–103713971	insertion	G[+1A]A
103715441–103715442	insertion	T[+1A]A
103724672–103724673	insertion	C[+1T]C
103724683–103724686	deletion	G[-2TT]T
103724688–103724689	insertion	A[+1C]C
103724693–103724694	insertion	C[+1A]A
103724699–103724701	deletion	C[-1T]T
103733727–103733734	deletion	G[-6AACCAC]A
103734107–103734145	deletion	T[-37AAAGAATACTCAAAACATCT ATCTTGGAGTACTAGCA]C
103737874–103737885	deletion	T[-10TGAACATCTC]A
103737931–103737941	deletion	G[-9CTATTAGGG]C
103737953–103737955	deletion	T[-1C]T
103739248–103739249	insertion	A[+8ATAAAACG]C
103740150–103740152	deletion	G[-1A]A
103740323–103740328	deletion	G[-4CCCA]A
103740336–103740337	insertion	G[+1A]A
103741390–103741394	deletion	A[-3GCT]G
103741576–103741579	deletion	A[-2CG]G
103741581–103741582	insertion	A[+1T]A
103741712–103741715	deletion	A[-2AC]A
103741716–103741718	deletion	T[-1A]G
103742086–103742093	deletion	C[-6TGTTGC]A
103742736–103742738	deletion	A[-1C]T

Table 4.3, continued.

103742947–103742948	insertion	G[+10TATTTACTTT]C
103742950–103742952	deletion	T[-1C]A
103743369–103743370	insertion	T[+5CAGTA]T
103744863–103744878	deletion	C[-14CATCCCGCAGAGGC]C
103744894–103744896	deletion	A[-1C]C
103745038–103745039	insertion	T[+1G]A
103745100–103745101	insertion	T[+7CCTTAAA]T
103791114–103791115	insertion	T[+1G]A
103794219–103794220	insertion	G[+2CA]G
103802161–103802162	insertion	T[+1A]A
103810712–103810718	deletion	A[-5TGAGT]A
103814090–103814093	deletion	T[-2AA]A
103824628–103824630	deletion	A[-1T]T
103826164–103826165	insertion	T[+1A]A
103828512–103828515	deletion	C[-2AG]C
103835567–103835568	insertion	T[+2TC]T
103836487–103836488	insertion	T[+1A]A
103837575–103837576	insertion	G[+2TT]G
103838414–103838415	insertion	T[+1A]A
103839090–103839099	deletion	C[-8CCCTCCCT]C
103839618–103839619	insertion	G[+1A]A
103842574–103842576	deletion	C[-1A]A
103843795–103843796	insertion	T[+2CC]C
103843802–103843807	deletion	A[-4CTTT]T
103843809–103843811	deletion	C[-1T]A
103843813–103843814	insertion	T[+1G]C
103843844–103843846	deletion	T[-1C]C
103843880–103843882	deletion	A[-1C]C
103843884–103843885	insertion	G[+1A]A
103844367–103844369	deletion	A[-1T]A
103844373–103844376	deletion	C[-2AG]T
103844377–103844381	deletion	G[-3TTC]A
103844387–103844388	insertion	T[+3GAA]C
103844393–103844394	insertion	T[+2TA]T
103845059–103845064	deletion	C[-4AGTT]A
103845222–103845224	deletion	G[-1A]A
103846591–103846594	deletion	G[-2GT]G
103848952–103848953	insertion	C[+1T]A
103851251–103851252	insertion	A[+1T]T
103851436–103851437	insertion	A[+1T]T
103891653–103891654	insertion	A[+3TTT]T
103891689–103891691	deletion	A[-1T]T
103892563–103892564	insertion	T[+9TAAAATCAC]G
103893035–103893036	insertion	T[+1A]G
103895394–103895395	insertion	T[+2AA]A

V. Future directions/Discussion

Although we have shown that the missplicing of *pteg* coincides with the *curly* phenotype, that fact does not in and of itself prove that the *pteg* deficiency causes the *curly* phenotype. Attempts by Isabelle to phenocopy the *curly* mutation using a morpholino oligonucleotide targeting the *pteg* gene have been inconclusive, as have attempts to rescue the mutant phenotype via mRNA injections (data not shown). Recently, however, a paper was published by the Grainger lab at the University of Virginia showing the accurate recapitulation of endogenous gene expression in *X. tropicalis* via BAC injection (Fish et al., 2011). Notably, they showed that injected BACs are amplified in the embryo, presumably allowing for nonmosaic expression that perdures through later stages, in contrast to injected mRNA, which eventually degrades (Harland and Misher, 1988). Rusconi and Schaffner showed that injected plasmids containing the rabbit β-globin gene, or the circularized gene itself could result in replication of the DNA, and expression of the globin gene, that could last for months in the frog (Rusconi and Schaffner, 1981). I have located four BACs whose end sequences suggest that they include the *pteg* locus (Figure 4.6), and we are hopeful that injection of one or more of these BACs will rescue the *curly* phenotype. We are currently refining our technique for BAC injection into 1-cell stage embryos from natural mating (as opposed to embryos from *in vitro* fertilization as reported, because our numbers of mutant carriers are limited) using the Pax6 BAC shown in Fish et al., and Isabelle is preparing the *pteg*-containing BACs for injection by incorporating a fluorescent reporter into the BACs via "recombineering".

In order to determine whether a rescue experiment has succeeded, we will genotype injected embryos using the marker 031D02 to determine which are homozygous *curly* and focus on those for comparison of phenotypes between *pteg* BAC injected and control injected embryos. This marker tends to amplify well and be clearly interpretable, and our mapping data have shown that it is over 99% accurate in predicting whether an embryo is a *curly* mutant (Table 4.1).

The mechanism by which *pteg* could cause the *curly* phenotype remains an open question. Lee et al. showed that in *X. laevis*, *pteg* knockdown caused a reduction in pronephric marker expression in the proximal tubules, concluding that Pteg is required for pronephric tubulogenesis (Lee et al., 2010). Pteg overexpression increased pronephric marker expression in the proximal tubules. The *X. laevis* and *X. tropicalis* forms of the protein share a high degree of homology, including the possession of a signal peptide, transmembrane domain, and putative PDZ-binding domain (Figure 4.4) (Lee et al., 2010). However, the targeted injection of the *pteg* MO performed by Lee et al. did not result in a *curly*-like phenotype, so we hypothesize that any effect on the kidney from a loss of proper *pteg* expression cannot fully explain the *curly* phenotype. Although Lee et al. only reported *pteg* expression in the kidney, Isabelle has shown that *pteg* is expressed in the heart region. She has also shown that the hearts of *curly* mutants do not develop properly, which is consistent with the ventral edema of *curly* mutants beginning around the heart. Furthermore, Nancy Hoo, an undergraduate in the Harland lab, showed that *curly* mutants display a dysregulated cell cycle that is followed by increased apoptosis in the late 30's stages, although we should

repeat this experiment with tadpoles from both the *blimpy* and *curly* lines. All in all, it is likely that a *pteg* deficiency has pleiotropic effects.

This project demonstrates the efficacy of mapping *X. tropicalis* mutations using microsatellite markers, and the potential for doing so even more expeditiously via next-generation sequencing. If the *curly* phenotype is indeed caused by a lack of functional *pteg* protein, it will indicate novel functions for this gene in heart morphogenesis and possibly in cell cycle regulation. Furthermore, the identification of the lesion that causes the missplicing of *pteg* may yield insight into the mechanism of splicing regulation.

VI. Additional materials and methods:

Mapping PCR (optimized by I. Philipp):

1 μL DNA
10 μL 10X PCR buffer (final concentrations 10 mM Tris-HCl pH 9, 50 mM KCl, 1.5 or 5 mM $MgCl_2$, 0.1% Triton)
0.4 μM each primer, final concentration
0.2 mM each dNTP, final concentration
0.5 μL 100X BSA (New England Biolabs)
1 μL 50X Taq polymerase
to 50 μL H_20

Mapping PCR program (D. Bhattacharya, personal communication):

94°C, 2 min
40 cycles of:
    94°C, 10 sec
    58°c or 54°C, 30 sec
    72°C, 30 sec
72°C, 5 min
10°C for ever

All markers used for mapping were resolved on 6% polyacrylamide gels made from 30% acrylamide solution with a 29:1 ratio of acrylamide to bis-acrylamide (Bio-Rad), in 0.56 M Tris-HCl pH 8.8.

A

	v.7 scaffold 4 coordinates	Length (bp)		
pteg transcript	103,779,159–103,788,639	9,480	bp 5' of *pteg*	bp 3' of *pteg*
OAAA127J20/ CH216-**127J20**	103,654,944–103,802,751	147,807	14,112	124,215
OAAA022D10/ CH216-**22D10**	103,655,577–103,802,865	147,288	14,226	123,582
OAAAB120K02/ ISB1-**120K2**	103,731,049–103,814,847	83,798	26,171	48,110
OAAAB040N06/ ISB1-**40N6**	103,747,131–103,819,613	72,482	30,974	32,028

B

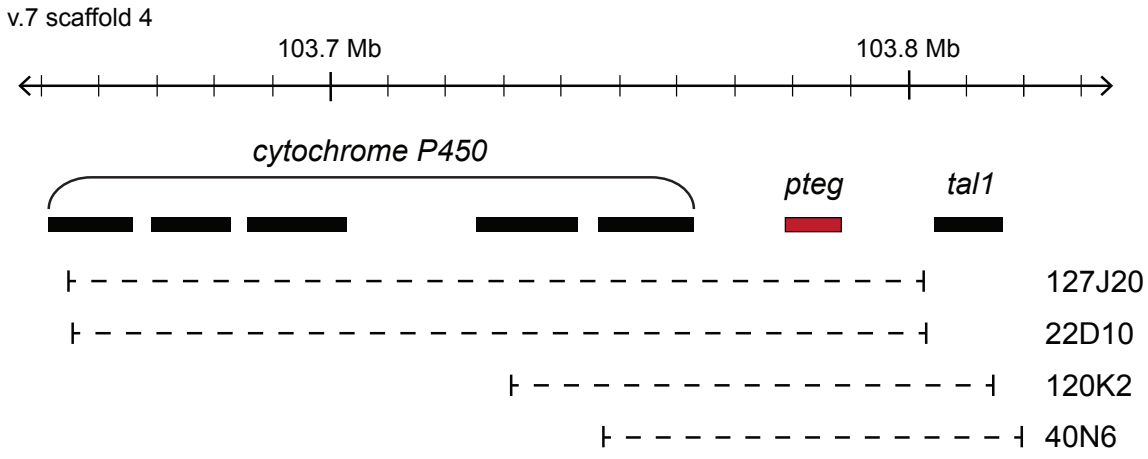


Figure 4.6. BACs that contain the pteg gene. (A) Coordinates of the *pteg* gene in *X. tropicalis* genome assembly version 7, and of BACs that contain *pteg*. (B) The four BACs contain different subsets of genes but all contain the full-length *pteg* gene. Only 40N6 is predicted to contain the full-length *tal1* gene.

References

Abu-Daya, A., Khokha, M. K. and Zimmerman, L. B. (2012) 'The Hitchhiker's guide to Xenopus genetics', *Genesis (New York, NY : 2000)*.

Abu-Daya, A., Sater, A. K., Wells, D. E., Mohun, T. J. and Zimmerman, L. B. (2009) 'Absence of heartbeat in the Xenopus tropicalis mutation muzak is caused by a nonsense mutation in cardiac myosin myh6', *Developmental Biology* 336(1): 20-9.

Altshuler, D., Pollara, V. J., Cowles, C. R., Van Etten, W. J., Baldwin, J., Linton, L. and Lander, E. S. (2000) 'An SNP map of the human genome generated by reduced representation shotgun sequencing', *Nature* 407(6803): 513-6.

Amaya, E., Musci, T. J. and Kirschner, M. W. (1991) 'Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in Xenopus embryos', *Cell* 66(2): 257-70.

Amaya, E., Offield, M. F. and Grainger, R. M. (1998) 'Frog genetics: Xenopus tropicalis jumps into the future', *Trends Genet* 14(7): 253-5.

Amaya, E., Stein, P. A., Musci, T. J. and Kirschner, M. W. (1993) 'FGF signalling in the early specification of mesoderm in Xenopus', *Development* 118(2): 477-87.

Avivi, A., Yayon, A. and Givol, D. (1993) 'A novel form of FGF receptor-3 using an alternative exon in the immunoglobulin domain III', *FEBS Lett* 330(3): 249-52.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A. and Johnson, E. A. (2008) 'Rapid SNP discovery and genetic mapping using sequenced RAD markers', *PLoS ONE* 3(10): e3376.

Beck, C. W. and Slack, J. M. (2001) 'An amphibian with ambition: a new role for Xenopus in the 21st century', *Genome Biol* 2(10): REVIEWS1029.

Bentley, D. R. (2006) 'Whole-genome re-sequencing', *Curr Opin Genet Dev* 16(6): 545-52.

Bentley, D. R. Balasubramanian, S. Swerdlow, H. P. Smith, G. P. Milton, J. Brown, C. G. Hall, K. P. Evers, D. J. Barnes, C. L. Bignell, H. R. et al. (2008) 'Accurate whole human genome sequencing using reversible terminator chemistry', *Nature* 456(7218): 53-9.

Bewick, A. J., Anderson, D. W. and Evans, B. J. (2010) 'EVOLUTION OF THE CLOSELY RELATED, SEX-RELATED GENES DM-W AND DMRT1 IN AFRICAN CLAWED FROGS (XENOPUS)', *Evolution; international journal of organic evolution*.

Bhattacharya, D. (2010) Developmental analysis and genetic mapping of the *grinch* mutation in *Xenopus tropicalis*. Undergraduate honors thesis: University of California, Berkeley.

Böttcher, R. T. and Niehrs, C. (2005) 'Fibroblast growth factor signaling during early vertebrate development', *Endocr Rev* 26(1): 63-77.

Carballada, R., Yasuo, H. and Lemaire, P. (2001) 'Phosphatidylinositol-3 kinase acts in parallel to the ERK MAP kinase in the FGF pathway during Xenopus mesoderm induction', *Development* 128(1): 35-44.

CHANG, C. Y. and WITSCHI, E. (1956) 'Genic control and hormonal reversal of sex differentiation in Xenopus', *Proc Soc Exp Biol Med* 93(1): 140-4.

Chapman, J. A., Ho, I., Sunkara, S., Luo, S., Schroth, G. P. and Rokhsar, D. S. (2011) 'Meraculous: de novo genome assembly with short paired-end reads', *PLoS ONE* 6(8): e23501.

Chellaiah, A., Yuan, W., Chellaiah, M. and Ornitz, D. M. (1999) 'Mapping ligand binding domains in chimeric fibroblast growth factor receptor molecules. Multiple regions determine ligand binding specificity', *J Biol Chem* 274(49): 34785-94.

Chellaiah, A. T., McEwen, D. G., Werner, S., Xu, J. and Ornitz, D. M. (1994) 'Fibroblast growth factor receptor (FGFR) 3. Alternative splicing in immunoglobulin-like domain III creates a receptor highly specific for acidic FGF/FGF-1', *J Biol Chem* 269(15): 11620-7.

Chung, H. A., Hyodo-Miura, J., Kitayama, A., Terasaka, C., Nagamune, T. and Ueno, N. (2004) 'Screening of FGF target genes in Xenopus by microarray: temporal dissection of the signalling pathway using a chemical inhibitor', *Genes Cells* 9(8): 749-61.

Craig, D. W., Pearson, J. V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J. J., Pawlowski, T. L., Laub, T., Nunn, G., Stephan, D. A. et al. (2008) 'Identification of genetic variants using bar-coded multiplexed sequencing', *Nat Methods* 5(10): 887-93.

Crew, F. A. (1939) 'Biological Tests for Diagnosis of Pregnancy', *Br Med J* 1(4084): 766-70.

Darnell, D. K., Stanislaw, S., Kaur, S. and Antin, P. B. (2010) 'Whole mount in situ hybridization detection of mRNAs using short LNA containing DNA oligonucleotide probes', *RNA* 16(3): 632-7.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M. and Blaxter, M. L. (2011) 'Genome-wide genetic marker discovery and genotyping using next-generation sequencing', *Nat Rev Genet* 12(7): 499-510.
Delaune, E., Lemaire, P. and Kodjabachian, L. (2005) 'Neural induction in Xenopus requires early FGF signalling in addition to BMP inhibition', *Development* 132(2): 299-310.

Dorey, K. and Amaya, E. (2010) 'FGF signalling: diverse roles during early vertebrate embryogenesis', *Development* 137(22): 3731-42.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. and Mitchell, S. E. (2011) 'A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species', *PLoS ONE* 6(5): e19379.

Eswarakumar, V. P., Lax, I. and Schlessinger, J. (2005) 'Cellular signaling by fibroblast growth factor receptors', *Cytokine Growth Factor Rev* 16(2): 139-49.

Evans, B. J., Kelley, D. B., Tinsley, R. C., Melnick, D. J. and Cannatella, D. C. (2004) 'A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution', *Mol Phylogenet Evol* 33(1): 197-213.

Ezaz, T., Stiglec, R., Veyrunes, F. and Marshall Graves, J. A. (2006) 'Relationships between vertebrate ZW and XY sex chromosome systems', *Curr Biol* 16(17): R736-43.

Fish, M. B., Nakayama, T. and Grainger, R. M. (2011) 'Simple, fast, tissue-specific bacterial artificial chromosome transgenesis in Xenopus', *Genesis (New York, NY : 2000)*.

Fisher, M. E., Isaacs, H. V. and Pownall, M. E. (2002) 'eFGF is required for activation of XmyoD expression in the myogenic cell lineage of Xenopus laevis', *Development* 129(6): 1307-15.

Fletcher, R. B., Baker, J. C. and Harland, R. M. (2006) 'FGF8 spliceforms mediate early mesoderm and posterior neural tissue formation in Xenopus', *Development* 133(9): 1703-14.

Fletcher, R. B. and Harland, R. M. (2008) 'The role of FGF signaling in the establishment and maintenance of mesodermal gene expression in Xenopus', *Dev Dyn* 237(5): 1243-54.

Frazzetto, G., Klingbeil, P. and Bouwmeester, T. (2002) 'Xenopus marginal coil (Xmc), a novel FGF inducible cytosolic coiled-coil protein regulating gastrulation movements', *Mech Dev* 113(1): 3-14.

Friesel, R. and Brown, S. A. (1992) 'Spatially restricted expression of fibroblast growth factor receptor-2 during Xenopus development', *Development* 116(4): 1051-8.

Geach, T. J. and Zimmerman, L. B. (2010) 'Paralysis and delayed Z-disc formation in the Xenopus tropicalis unc45b mutant dicky ticker', *BMC Dev Biol* 10: 75.
Gilchrist, M. J., Zorn, A. M., Voigt, J., Smith, J. C., Papalopulu, N. and Amaya, E. (2004) 'Defining a large set of full-length clones from a Xenopus tropicalis EST project', *Developmental Biology* 271(2): 498-516.

Givol, D. and Yayon, A. (1992) 'Complexity of FGF receptors: genetic basis for structural diversity and functional specificity', *FASEB J* 6(15): 3362-9.

Golub, R., Adelman, Z., Clementi, J., Weiss, R., Bonasera, J. and Servetnick, M. (2000) 'Evolutionarily conserved and divergent expression of members of the FGF receptor family

among vertebrate embryos, as revealed by FGFR expression patterns in Xenopus', *Dev Genes Evol* 210(7): 345-57.

Gospodarowicz, D. (1974) 'Localisation of a fibroblast growth factor and its effect alone and with hydrocortisone on 3T3 cell growth', *Nature* 249(453): 123-7.

Grammer, T. C., Khokha, M. K., Lane, M. A., Lam, K. and Harland, R. M. (2005) 'Identification of mutants in inbred Xenopus tropicalis', *Mech Dev* 122(3): 263-72.

Gurdon, J. B. and Hopwood, N. (2000) 'The introduction of Xenopus laevis into developmental biology: of empire, pregnancy testing and ribosomal genes', *Int J Dev Biol* 44(1): 43-50.

Harland, R. and Misher, L. (1988) 'Stability of RNA in developing Xenopus embryos and identification of a destabilizing sequence in TFIIIA messenger RNA', *Development* 102(4): 837-52.

Hayes, T. B. (1998) 'Sex determination and primary sex differentiation in amphibians: genetic and developmental mechanisms', *J Exp Zool* 281(5): 373-99.

Hayes, T. B., Collins, A., Lee, M., Mendoza, M., Noriega, N., Stuart, A. A. and Vonk, A. (2002) 'Hermaphroditic, demasculinized frogs after exposure to the herbicide atrazine at low ecologically relevant doses', *Proc Natl Acad Sci USA* 99(8): 5476-80.

Heasman, J. (2002) 'Morpholino oligos: making sense of antisense?', *Developmental Biology* 243(2): 209-14.

Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N. H., Shu, S., Taher, L. et al. (2010) 'The genome of the Western clawed frog Xenopus tropicalis', *Science* 328(5978): 633-6.

Hellsten, U., Khokha, M. K., Grammer, T. C., Harland, R. M., Richardson, P. and Rokhsar, D. S. (2007) 'Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog Xenopus laevis', *BMC Biol* 5: 31.

Isaacs, H. V., Pownall, M. E. and Slack, J. M. (1994) 'eFGF regulates Xbra expression during Xenopus gastrulation', *EMBO J* 13(19): 4469-81.
Isaacs, H. V., Tannahill, D. and Slack, J. M. (1992) 'Expression of a novel FGF in the Xenopus embryo. A new candidate inducing factor for mesoderm formation and anteroposterior specification', *Development* 114(3): 711-20.

Itoh, N. and Ornitz, D. M. (2004) 'Evolution of the Fgf and Fgfr gene families', *Trends Genet* 20(11): 563-9.

Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M. et al. (2011) 'Mouse genomic variation and its effect on phenotypes and gene regulation', *Nature* 477(7364): 289-94.

Khokha, M. K. (2012) 'Xenopus white papers and resources: folding functional genomics and genetics into the frog', *Genesis (New York, NY : 2000)* 50(3): 133-42.

Khokha, M. K., Chung, C., Bustamante, E. L., Gaw, L. W. K., Trott, K. A., Yeh, J., Lim, N., Lin, J. C. Y., Taverner, N., Amaya, E. et al. (2002) 'Techniques and probes for the study of Xenopus tropicalis development', *Dev Dyn* 225(4): 499-510.

Khokha, M. K., Krylov, V., Reilly, M. J., Gall, J. G., Bhattacharya, D., Cheung, C. Y. J., Kaufman, S., Lam, D. K., Macha, J., Ngo, C. et al. (2009) 'Rapid gynogenetic mapping of Xenopus tropicalis mutations to chromosomes', *Dev Dyn* 238(6): 1398-46.

Kimelman, D., Abraham, J. A., Haaparanta, T., Palisi, T. M. and Kirschner, M. W. (1988) 'The presence of fibroblast growth factor in the frog egg: its role as a natural mesoderm inducer', *Science* 242(4881): 1053-6.

Kimelman, D. and Kirschner, M. (1987) 'Synergistic induction of mesoderm by FGF and TGF-beta and the identification of an mRNA coding for FGF in the early Xenopus embryo', *Cell* 51(5): 869-77.

Kornblihtt, A. R. (2005) 'Promoter usage and alternative splicing', *Curr Opin Cell Biol* 17(3): 262-8.

Lamb, T. M. and Harland, R. M. (1995) 'Fibroblast growth factor is a direct neural inducer, which combined with noggin generates anterior-posterior neural pattern', *Development* 121(11): 3627-36.

Lane, M. A., Mustafa K. Khokha, Connie Ng, Richard M. Harland, Timothy C. Grammer. Genetic identification of an axanthic pigment variant in *Xenopus tropicalis*. In preparation.

Lea, R., Papalopulu, N., Amaya, E. and Dorey, K. (2009) 'Temporal and spatial expression of FGF ligands and receptors during Xenopusdevelopment', *Dev. Dyn.* 238(6): 1467-1479.
Lee, S. J., Kim, S., Choi, S.-C. and Han, J.-K. (2010) 'XPteg (Xenopus proximal tubules-expressed gene) is essential for pronephric mesoderm specification and tubulogenesis', *Mech Dev* 127(1-2): 49-61.

Lennon, N. J., Lintner, R. E., Anderson, S., Alvarez, P., Barry, A., Brockman, W., Daza, R., Erlich, R. L., Giannoukos, G., Green, L. et al. (2010) 'A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454', *Genome Biol* 11(2): R15.

Lombardo, A., Isaacs, H. V. and Slack, J. M. (1998) 'Expression and functions of FGF-3 in Xenopus development', *Int J Dev Biol* 42(8): 1101-7.

Love, N. R., Thuret, R., Chen, Y., Ishibashi, S., Sabherwal, N., Paredes, R., Alves-Silva, J., Dorey, K., Noble, A. M., Guille, M. J. et al. (2011) 'pTransgenesis: a cross-species, modular transgenesis resource', *Development* 138(24): 5451-8.

Lund, E., Sheets, M. D., Imboden, S. B. and Dahlberg, J. E. (2011) 'Limiting Ago protein restricts RNAi and microRNA biogenesis during early development in Xenopus laevis', *Genes Dev* 25(11): 1121-31.

MacNicol, A. M., Muslin, A. J. and Williams, L. T. (1993) 'Raf-1 kinase is essential for early Xenopus development and mediates the induction of mesoderm by FGF', *Cell* 73(3): 571-83.

Metzker, M. L. (2010) 'Sequencing technologies - the next generation', *Nat Rev Genet* 11(1): 31-46.

Miki, T., Bottaro, D. P., Fleming, T. P., Smith, C. L., Burgess, W. H., Chan, A. M. and Aaronson, S. A. (1992) 'Determination of ligand-binding specificity by alternative splicing: two distinct growth factor receptors encoded by a single gene', *Proc Natl Acad Sci USA* 89(1): 246-50.

Monsoro-Burq, A.-H. (2003) 'Neural crest induction by paraxial mesoderm in Xenopus embryos requires FGF signals', *Development* 130(>14): 3111-3124.

Nasevicius, A. and Ekker, S. C. (2000) 'Effective targeted gene 'knockdown' in zebrafish', *Nat Genet* 26(2): 216-20.

Nieuwkoop, P. D., Jacob Faber (1994) *Normal Table of Xenopus laevis (Daudin)*, New York: Garland Publishing.

Nutt, S. L., Dingwell, K. S., Holt, C. E. and Amaya, E. (2001) 'Xenopus Sprouty2 inhibits FGF-mediated gastrulation movements but does not affect mesoderm induction and patterning', *Genes Dev* 15(9): 1152-66.

Odenthal, J., Rossnagel, K., Haffter, P., Kelsh, R. N., Vogelsang, E., Brand, M., van Eeden, F. J., Furutani-Seiki, M., Granato, M., Hammerschmidt, M. et al. (1996) 'Mutations affecting xanthophore pigmentation in the zebrafish, Danio rerio', *Development* 123: 391-8.

Olmstead, A. W., Lindberg-Livingston, A. and Degitz, S. J. (2010) 'Genotyping sex in the amphibian, Xenopus (Silurana) tropicalis, for endocrine disruptor bioassays', *Aquat Toxicol* 98(1): 60-6.

Ornitz, D. M. and Marie, P. J. (2002) 'FGF signaling pathways in endochondral and intramembranous bone development and human genetic disease', *Genes Dev* 16(12): 1446-65.

Ornitz, D. M., Xu, J., Colvin, J. S., McEwen, D. G., MacArthur, C. A., Coulier, F., Gao, G. and Goldfarb, M. (1996) 'Receptor specificity of the fibroblast growth factor family', *J Biol Chem* 271(25): 15292-7.

Orr-Urtreger, A., Bedford, M. T., Burakova, T., Arman, E., Zimmer, Y., Yayon, A., Givol, D. and Lonai, P. (1993) 'Developmental localization of the splicing alternatives of fibroblast growth factor receptor-2 (FGFR2)', *Developmental Biology* 158(2): 475-86.

Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M. and Fire, A. Z. (2007) 'A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing', *Nucleic Acids Res* 35(19): e130.

Parichy, D. M. and Turner, J. M. (2003) 'Temporal and cellular requirements for Fms signaling during zebrafish adult pigment pattern development', *Development* 130(5): 817-33.

Plotnikov, A. N., Hubbard, S. R., Schlessinger, J. and Mohammadi, M. (2000) 'Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity', *Cell* 101(4): 413-24.

Pope, A. P., Liu, C., Sater, A. K. and Servetnick, M. (2010) 'FGFR3 expression in Xenopus laevis', *Gene Expr Patterns* 10(2-3): 87-92.

Quail, M. A., Kozarewa, I., Smith, F., Scally, A., Stephens, P. J., Durbin, R., Swerdlow, H. and Turner, D. J. (2008) 'A large genome center's improvements to the Illumina sequencing system', *Nat Methods* 5(12): 1005-10.

Ribisi, S., Mariani, F. V., Aamar, E., Lamb, T. M., Frank, D. and Harland, R. M. (2000) 'Ras-mediated FGF signaling is required for the formation of posterior but not anterior neural tissue in Xenopus laevis', *Developmental Biology* 227(1): 183-96.

Rusconi, S. and Schaffner, W. (1981) 'Transformation of frog embryos with a rabbit beta-globin gene', *Proc Natl Acad Sci USA* 78(8): 5051-5.
Schlessinger, J., Plotnikov, A. N., Ibrahimi, O. A., Eliseenkova, A. V., Yeh, B. K., Yayon, A., Linhardt, R. J. and Mohammadi, M. (2000) 'Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization', *Mol Cell* 6(3): 743-50.

Shi, W. and Levine, M. (2008) 'Ephrin signaling establishes asymmetric cell fates in an endomesoderm lineage of the Ciona embryo', *Development* 135(5): 931-40.

Sive, H. L., Robert M. Grainger, Richard M. Harland (2000) *Early Development of Xenopus laevis: A Laboratory Manual*, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Slack, J. M., Darlington, B. G., Heath, J. K. and Godsave, S. F. (1987) 'Mesoderm induction in early Xenopus embryos by heparin-binding growth factors', *Nature* 326(6109): 197-200.

Slack, J. M. W., Lin, G. and Chen, Y. (2008) 'The Xenopus tadpole: a new model for regeneration research', *Cell Mol Life Sci* 65(1): 54-63.

Song, J. and Slack, J. M. (1996) 'XFGF-9: a new fibroblast growth factor from Xenopus embryos', *Dev Dyn* 206(4): 427-36.

Standley, H. J., Zorn, A. M. and Gurdon, J. B. (2001) 'eFGF and its mode of action in the community effect during Xenopus myogenesis', *Development* 128(8): 1347-57.

Summerton, J. and Weller, D. (1997) 'Morpholino antisense oligomers: design, preparation, and properties', *Antisense Nucleic Acid Drug Dev* 7(3): 187-95.

Torpey, N., Wylie, C. C. and Heasman, J. (1992) 'Function of maternal cytokeratin in Xenopus development', *Nature* 357(6377): 413-5.

Trivedi, T. (2009) Genetic mapping of early developmental mutation *curly* in *Xenopus tropicalis*. Undergraduate honors thesis: University of California, Berkeley.

Tymowska, J. and Fischberg, M. (1973) 'Chromosome complements of the genus Xenopus', *Chromosoma* 44(3): 335-42.

Umbhauer, M., Marshall, C. J., Mason, C. S., Old, R. W. and Smith, J. C. (1995) 'Mesoderm induction in Xenopus caused by activation of MAP kinase', *Nature* 376(6535): 58-62.

Umbhauer, M., Penzo-Méndez, A., Clavilier, L., Boucaut, J. and Riou, J. (2000) 'Signaling specificities of fibroblast growth factor receptors in early Xenopus embryo', *J Cell Sci* 113 (Pt 16): 2865-75.

Wang, J. K., Gao, G. and Goldfarb, M. (1994) 'Fibroblast growth factor receptors have different signaling and mitogenic potentials', *Molecular and Cellular Biology* 14(1): 181-8.

Wells, D. E., Gutierrez, L., Xu, Z., Krylov, V., Macha, J., Blankenburg, K. P., Hitchens, M., Bellot, L. J., Spivey, M., Stemple, D. L. et al. (2011) 'A genetic map of Xenopus tropicalis', *Developmental Biology*.

Xu, Z., Gutierrez, L., Hitchens, M., Scherer, S., Sater, A. K. and Wells, D. E. (2008) 'Distribution of polymorphic and non-polymorphic microsatellite repeats in Xenopus tropicalis', *Bioinform Biol Insights* 2: 157-69.

Yalcin, B., Wong, K., Agam, A., Goodson, M., Keane, T. M., Gan, X., Nellåker, C., Goodstadt, L., Nicod, J., Bhomra, A. et al. (2011) 'Sequence-based characterization of structural variation in the mouse genome', *Nature* 477(7364): 326-9.

Yamagishi, M. and Okamoto, H. (2010) 'Competition for ligands between FGFR1 and FGFR4 regulates Xenopus neural development', *Int J Dev Biol* 54(1): 93-104.

Yan, G., Fukabori, Y., McBride, G., Nikolaropolous, S. and McKeehan, W. L. (1993) 'Exon switching and activation of stromal and embryonic fibroblast growth factor (FGF)-FGF receptor genes in prostate epithelial cells accompany stromal independence and malignancy', *Molecular and Cellular Biology* 13(8): 4513-22.

Yayon, A., Zimmer, Y., Shen, G. H., Avivi, A., Yarden, Y. and Givol, D. (1992) 'A confined variable region confers ligand specificity on fibroblast growth factor receptors: implications for the origin of the immunoglobulin fold', *EMBO J* 11(5): 1885-90.

Yoshimoto, S., Ikeda, N., Izutsu, Y., Shiba, T., Takamatsu, N. and Ito, M. (2010) 'Opposite roles of DMRT1 and its W-linked paralogue, DM-W, in sexual dimorphism of Xenopus laevis: implications of a ZZ/ZW-type sex-determining system', *Development* 137(15): 2519-26.

Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., Matsuda, Y., Takamatsu, N., Shiba, T. and Ito, M. (2008) 'A W-linked DM-domain gene, DM-W, participates in primary ovary development in Xenopus laevis', *Proc Natl Acad Sci USA* 105(7): 2469-74.

Young, J. J., Cherone, J. M., Doyon, Y., Ankoudinova, I., Faraji, F. M., Lee, A. H., Ngo, C., Guschin, D. Y., Paschon, D. E., Miller, J. C. et al. (2011) 'Efficient targeted gene disruption in the soma and germ line of the frog Xenopus tropicalis using engineered zinc-finger nucleases', *Proc Natl Acad Sci USA* 108(17): 7052-7.

Zhang, X., Ibrahimi, O. A., Olsen, S. K., Umemori, H., Mohammadi, M. and Ornitz, D. M. (2006) 'Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family', *J Biol Chem* 281(23): 15694-700.

Appendix

Table A1. PCR primers used in this work.

Primer name	Primer sequence	Purpose
pteg-ATG-alpha-F	5' ATGTTATCCCTGCAACATTTGC 3'	Used for *pteg* RT-PCR: they bind in exon 1 and 4, respectively.
pteg-TAA-omega-R	5' TTACATAGCTGTAGTACGGAC 3'	
s350-936874-F	5' AGGATGACAATTCACGAGTCC 3'	Designed by I. Philipp; used for SSLP mapping PCR.
s350-936874-R	5' CATCACTGTCAGTACCATCAAGG 3'	
Sc4_107174736-F	5' GAACTTCGGCCTATAGAATTTGC 3'	Used for SSLP mapping PCR.
Sc4_107174736-R	5' TGCCTTGAACAAGATAACAATGG 3'	
odclexon6_4-23	5' TGTTCTGCGCATAGCAACTG 3'	Loading control for RT-PCR in *X. tropicalis* .
odcExn6-7_203-183	5' ACATCGTGCATCTGAGACAGC 3'	
tropEF1a_exon5For	5' CCCTGCTGGAAGCTCTTGAC 3'	Loading control for RT-PCR. Usable for *X. tropicalis* and *X. laevis* .
tropEF1a_exon6Rev	5' GGACACCAGTCTCCACACGA 3'	
FGFR1exon7-8_36-55	5' CTCCCAGCGAATACGAGTGT 3'	Assay expression of Fgfr1 IIIb isoform via RT-PCR
FGFR1exon7-8_228-208	5' GGTCAAAACTTCTGCGTCTGA 3'	
FR1exon7_109-128	5' CTCACATCCAATGGCTCAGG 3'	Assay expression of Fgfr1 IIIc isoform via RT-PCR
FR1exon7-9_292-273	5' AGTTAGCGGCCAAGCAGGTA 3'	
FR2Exon6_113-132	5' CATCCGCTGGGTGAGATACA 3'	Assay expression of Fgfr2 IIIb isoform via RT-PCR
FR2exon6-7_265-246	5' ATTCCCCTGCGTCCTCTTCT 3'	
FR2exon6_71-90	5' CGCAGAGTTTGTCTGCAAGG 3'	Assay expression of Fgfr2 IIIc isoform via RT-PCR
FR2exon6-8_270-251	5' TTCCCCAGCATCCTCAAAAG 3'	
FGFR3exon6_100-119	5' ATGCCCAGCCTCATATTGAC 3'	Assay expression of Fgfr3 IIIb isoform via RT-PCR
FR3exon7_1622-1603	5' CAAAAGGACGCCTCAGCTAC 3'	
FR3exon6-8_299-279	5' CCAATAGAATTCCCAGCCAGA 3'	With FR3exon6_100-119, used to assay expression of fgfr3 IIIc isoform via RT-PCR
FR4exon2_129-146	5' GGGAAGATTCGCATGGTG 3'	Assay efficacy of MOSAFR4_I2E3 MO Designed to work in *X. laevis* and *X. tropicalis* .
FR4exn2-3_301-282	5' TACGGCCATCCTCATCATCT 3'	
FR3exon4_94-112	5' GGAATCCCACCCCTACCAT 3'	Assay efficacy of XtFGFR3(I3E4) MO
FR3exn4-5_282-262	5' TTGATAGGTTTGACGGATGCT 3'	
FR1exon4_46-67	5' CATCCTCTGAGGAGAAAGCTTC 3'	Assay efficacy of XtFGFR1(I3E4) MO
FR1exn4-5_250-233	5' ATCCACCAATGCGCTGAT 3'	

Table A2. Plasmids used in this work.

Plasmid number	Name	Contains	Used for	Cut with	Transcribe with	Notes
trop #337	XtFGFR1-107-2	fgfr1	mRNA	NotI	SP6	
#1824		fgfr2	mRNA	NotI	SP7	
trop #346	XtFGFR3-107	fgfr3	mRNA	NotI	SP8	
#1820	XFGFR-4a	fgfr4a	mRNA	AvaI	SP6	
#1822	XFGFR4b	fgfr4b	mRNA	SmaI	SP6	
#2161	XLFGF8b CDS-CS8	fgf8b	in situ probe	EcoRI	T7	
			mRNA	AscI	SP6	
#170	pXBra	X. laevis t	in situ probe	XhoI	SP6	
trop#30	t	X. tropicalis t	in situ probe	EcoRI	T7	
#98	XMyf5-2	myf5	in situ probe	BglII	SP6	
#943	xvent-2/pBS	vent2	in situ probe	SalI	T7	
#164	pG500	goosecoid	in situ probe	BamHI	T3	
#688	pBSsk- chordin	chordin	in situ probe	EcoRI	T7	
#1751	Bix4	bix4	in situ probe	BamHI	T7	
#1099	pBSXSOX17B	sox17b	in situ probe	EcoRI	T7	
#671	CS2+nbetagal	nuclear-localized LacZ	mRNA	NotI	SP6	
#341	XtFR1isp7-1	fgfr1 3'UTR	in situ probe	StyI	T7	bp 1-1121 of 3'UTR
		fgfr2 3'UTR	in situ probe	SalI	T7	bp 1-450 of 3'UTR, in CS108: 5' end SalI; 3' end XhoI
	FR2exon7-CS108	fgfr2 IIIb exon	in situ probe	EcoRI	T7	5' end Eco RI, 3' end XhoI
	FR2exon8-CS108	fgfr2 IIIc exon	in situ probe	EcoRI	T7	5' end Eco RI, 3' end XhoI
	FR1exon8-CS108	fgfr1 IIIb exon	in situ probe	EcoRI	T7	5' end Eco RI, 3' end XhoI
	FR1exon9-CS108	fgfr1 IIIc exon	in situ probe	EcoRI	T7	5' end Eco RI, 3' end XhoI
	laevisFR4a-CS108	laevisFR4a, no UTR	mRNA	AscI	SP6	5' end NotI, 3' end XhoI
	tropFR3-CS108	tropFR3, no UTR	mRNA	AscI	SP6	5' end NotI, 3' end XhoI
		C. intestinalis torso-fgfr	mRNA			
		X. laevis torso-fgfr4	mRNA	SacI	SP6	