2009

Evolutionary Analysis of the Glutamic Acid Decarboxylase 67 Immediate Upstream Region in African Clawed Frogs

Jonathan Lomax Boyd
College of William & Mary - Arts & Sciences

Evolutionary Analysis of the Glutamic Acid Decarboxylase 67 Immediate Upstream Region in African Clawed Frogs

Jonathan Lomax Boyd

Virginia Beach, Virginia

Bachelor of Science, University of Richmond, 2006

A Thesis presented to the Graduate Faculty
of the College of William and Mary in Candidacy for the Degree of
Master of Science

Department of Biology

The College of William and Mary
August, 2009

APPROVAL PAGE

This Thesis is submitted in partial fulfillment of
the requirements for the degree of

Master of Science

Jonathan ~~Lomax~~ Boyd

Approved by the Committee, July, 2009

Committee Chair
Chancellor Professor Margaret Saha, Biology
The College of William and Mary

Hamilton Professor Lizabeth Allison, Biology
The College of William and Mary

Marjorie S. Curtis Term Distinguished Associate Professor George Gilchrist, Biology
The College of William and Mary

COMPLIANCE PAGE

Research approved by

Institutional Animal Care and Use Committee

Institutional Biosafety Committee

Protocol number(s): 1ACUC-2007-11-13-5001

1BC-2007-02-15-4594

Date(s) of approval: 11/04/2008

02/17/2009

ABSTRACT PAGE

Changes in the expression pattern of genes during development has been shown to underlie dramatic evolutionary changes. The relationship, however, between sequence variation within non-coding regulatory regions of DNA and transcriptional output remains poorly understood in the conext of molecular evolution. We report an evolutionary analysis of the immediate upstream region of xGAD67 in Xenopus laevis laevis . xGAD67 is the rate limiting enzyme primarily responsible for the acquisition of the inhibitory GABAergic neurotransmitter phenotype in the developing vertebrate central nervous system. Expression of GAD67 is part of the terminal differentiation of neural cells that use GABA (gamma-aminobutyric acid) as their neurotransmitter. The expression pattern of GAD67 during embryonic development has been shown to be highly similar among vertebrates where such studies are available, suggesing that the regulatory elements coordinating expression may be under selective pressure. Using a natural population of X. l. laeivs frogs and several closely related species we attempted to detect the action of selection with a variety of test statistics on the immediate upstream region of xGAD67. Analysis of the polymorphism frequency spectrum suggests the action of purifying selection around position -200 regardless of the Xenopus outgroup used in the analysis. The action of positive selection was also detected around position -900 and -1200 for X. l. sudanensis and X. amieti, respectively. Inferences of selection based upon patterns of polymorphim may provided an alternative approach to identifying uncharacterized regulatory elements.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Many people are responsible, some in small and subtle ways while others in more significant and obvious ways, for making the preparation of this manuscript feasible. On the familial side, my brothers, Joe and Justin, and parents unquestionably supported my eggheaded and financially irrational decision to return to science. Without that support none of the following pages could have been written and my future in science would have been remarkably different. On the professional side, Margaret Saha and Ralph Haygood were equal, yet disparate, sources of encouragement and never hesitant or frugal with scientific advise. I am indebted to my enablers, listed above and others, for the unpredictability of future results.

INTRODUCTION

Development of the vertebrate central nervous system is driven by conserved genetic networks from tissue specification to cell differentiation. The intricate structure of neural networks and the behavioral repertoire they generate have led to intense interest in uncovering the underlying genetics of neural development and the mechanisms by which these networks have evolved. Driven by differential gene expression, a comprehensive understanding of development will require knowledge of how genes important to neural differentiation are regulated at all organizational levels. Our focus is on the evolutionary pressures acting on the immediate upstream region of glutamic acid decarboxylase 67 (*xGAD67*), an enzyme that confers the GABAergic phenotype in the developing central nervous system (CNS) of vertebrates. We investigated the amount and pattern of variation in the upstream region of *xGAD67* in a natural population of *X. laevis laevis* individuals and several closely related species. The noncoding 5' flanking region of many genes has been shown to harbor functional, experimentally verified *cis*-regulatory elements (Wray et al., 2003). Moreover, these 5' proximal promoter elements have been proposed to constitute a major mechanism of vertebrate brain evolution (Britten and Davidson, 1971; Haygood et al., 2007; King and Wilson, 1975).

Neural Development Overview

Neural development in vertebrates invokes a succession of cell fate restrictions, which controls the process by which undifferentiated progenitor cells become fully functioning neurons. Neuronal precursor cells can be identified in the blastula where

pluripotent cells become segregated from presumptive epidermal precursors. Induction of neural tissue is mediated primarily through vertical signaling from the dorsal meso-endoderm to the ectoderm. The inductive potential of the dorsal mesoderm to direct neurulation and axis formation was demonstrated by the well known Spemann and Mangold organizer experiments (Bouwmeester, 2001), and later shown to result from β-catenin and Vg1/VegtT coexpression and decreased bone morphogenetic protein (BMP) inhibition. Early neurulation is characterized by the formation and shaping of the neural plate. The plate then thickens during mid-neurula and bends to shape the neural groove, which will invaginate to form the neutral tube. By late neurulation, localized expression of *N-cadherin* and *N-CAM* has initiated separation of the neural tube from the overlaying ectoderm. Complete closure requires, at least in mammals, *Pax3, Sonic hedgehog,* and *openbrain* expression (Milunsky et al., 1989).

Following neural tube closure, the neural tube becomes subdivided into major regions of the CNS along the anterior-posterior (A/P) axis. Expression of *Hox* genes defines gross anterior and posterior regions while the ectoderm differentiates into epidermal and neural fated tissue based on the specific activation or inhibition of BMP and Wnt proteins. Full inhibition of both BMP and Wnt results in the most anterior structures of the CNS, while inhibition of BMP and activation of Wnt gives rise to the spinal cord. The anterior region develops three characteristic vesicles that will become the forebrain, midbrain, and hindbrain. The spinal cord becomes morphologically distinct from the anterior CNS by an occlusion of cerebrospinal fluid constricting the neural tube at the base of the hindbrain (Desmond, 1982; Desmond and Levitan, 2002).

The dorsal-ventral (D/V) axis of the neural tube is polarized concomitantly with the A/P axis. The features of the ventral neural tube are induced by a gradient of Sonic hedgehog secretion emanating from the notochord. The most ventral region of the neural tube, where the concentration of Sonic hedgehog is highest, becomes the floor plate. The dorsal region of the neural tube is induced by secretion of TGF-β proteins (e.g. BMP4, BMP7, activin) from the overlying epidermis. The roof plate forms where the concentration of TGF-β proteins are highest and secretes secondary signaling molecules. Different layers of interneuron precursors are specified along the D/V axis by exposure to varying concentrations of Sonic hedgehog, TGF-β, and induced downstream regulators (e.g. *Pax7*, *Nkx6.1*, *Nkx2.2*, and *Pax6*) (Ericson et al., 1996).

As master regulatory genes define the A/P and D/V axis, a signaling cascade of neurogenic transcription factors, such as *SoxD*, *Neurogenin*, and *NeuroD* coordinate the downstream expression of terminal differentiation genes characteristic of neurons. For example, the expression of neural specific cytoskeletal structures, voltage-gated ion-channels, and biosynthetic enzymes required for production of some neurotransmitters are required for a fully differentiated neuron. In general, neurons acquire either inhibitory or excitatory neurotransmitter phenotypes whereby they communicate with neighboring cells through the release of small molecules (i.e. neurotransmitters) that cause inhibitory or excitatory post-synaptic responses.

Glutamic Acid Decarboxylase Expression

γ-Amino butyric acid (GABA) is the predominant inhibitory neurotransmitter utilized in both the vertebrate and invertebrate CNS (Watanabe et al., 2002). Acquisition of the GABAergic phenotype generally requires expression of the biosynthetic enzyme glutamic acid decarboxylase (GAD), which catalyzes the decarboxylation of glutamate to form GABA. Two predominant GAD isoforms, *GAD67* and *GAD65*, have been identified in many vertebrates (Arata et al., 2008; Bu et al., 1992; Erlander et al., 1991; Martyniuk et al., 2007; Trabucchi et al., 2008), and are expressed during early development and adulthood, respectively (Pinal and Tobin, 1998; Popp et al., 2009).

GAD65 is primarily found in the terminal bulbs of mature GABAergic neurons and is used for neurotransmitter mediated synaptic communication in adults (Martin and Rimvall, 1993). On the other hand, *GAD67* is located in both the cytoplasm of the soma and terminal bulbs of presumptive GABAergic neurons (Kaufman et al., 1991). While *GAD67* produced GABA may be used in neurotransmission, its roles as a trophic factor in synaptogenesis and providing protection during oxidative stress has been demonstrated (Allain et al., 2006; Lamigeon et al., 2001; Pinal and Tobin, 1998; Popp et al., 2009; Waagepetersen et al., 1999). *GAD67* is also known to promote cell proliferation, migration, differentiation, and cell death (Owens and Kriegstein, 2002). Both isoforms of GAD are differentially regulated (Popp et al., 2009; Somogyi et al., 1995); however, two variants of GAD67, GAD44 and GAD25, have been identified using Western blot analysis (Behar et al., 1993). Both GAD67 and GAD65 are also known to undergo post-translational modification (Buddhala et al., 2009; Wei and Wu, 2008).

The gross morphological features of the vertebrate CNS are remarkably similar between mammalian and non-mammalian animals. Master regulatory genes that specify and differentiate the neural tube, such as *Pax6*, *Dlx1/2*, *Emx1/2*, and *Tbr1* are similarly expressed. Regional expression of neurogenic and proneural genes in pallial and subpallial regions are comparable despite topological differences in telencephalon development, namely evagination in mammals and eversion is teleosts (Mueller et al., 2006; Wullimann and Mueller, 2004). The location of GABAergic neurons is similarly concentrated in the spinal cord, hindbrain, midbrain, and forebrain (telencephalon) of vertebrates (Guirado et al., 1999; Katarova et al., 2000; Marin et al., 1998; Medina and Reiner, 1995; Reiner et al., 1998).

The similarity of cellular phenotypes between deeply diverged vertebrates suggests the deployment of an equally conserved regulatory network. Changes within regulatory networks, such as mutations that alter the regulation or structure of homeotic genes, reveal the delicacy of signaling systems that define major morphological structures during development. We suggest that the upstream *cis*-regulatory elements (CREs) directing *GAD67* expression are under measurable functional constraint. Although many of the *cis*-regulatory elements coordinating embryonic *GAD67* expression have not been identified, the most likely location of putative regulatory elements is immediately upstream of the translational start site. The conserved expression of *GAD67* across distantly related vertebrate species establishes an *a priori* expectation that the underlying regulatory elements should be under purifying selection.

Evolution of Cis-regulatory Elements

Changes in noncoding, regulatory regions of DNA have long been suspected as a contributor to evolutionary innovation (Britten and Davidson, 1971; Jacob and Monod, 1961; King and Wilson, 1975). Molecular evolutionary analysis of gene expression centers on understanding the significance of sequence variation in *trans* acting factors and *cis* acting elements. The former is confined to the DNA sequence of transcription factor (TF) proteins that act in *trans* to regulate transcriptional activation. *Cis* acting regulatory elements are short sequences of DNA, referred to as transcription factor binding sites (TFBS), which are located adjacent to the protein-coding region being regulated. More broadly, however, there is contention whether the predominant mechanism of evolutionary innovation emerges from mutations in protein coding or non-coding, presumably regulatory, regions of DNA.

The rapid expansion of complete genome sequences for many species, including the human genome and many other closely related primates, has provided significant data to the *cis* verse *trans* debate. For example, the hypothesis by King and Wilson that *cis* regulatory mutations were the major mechanism behind human specific evolution has been supplemented by complete genome sequences. While they were correct that chimpanzees and humans are highly similar at the DNA level, at least 80% of the protein coding genes contain at lease one amino acid change (Glazko et al., 2005). Moreover, several studies have implicated changes in protein coding regions, such as transcription factors and sensory perception and immune system genes, that have undergone rapid evolution via natural selection along the human lineage (Bustamante et al., 2005; Clark et al., 2003; Dorus et al., 2004; Nielsen et al., 2005).

In addition to the genome-level surveys, which attempt to summarize the pattern of sequence divergence between species, investigation at individual genes has provided insight into the *cis* verse *trans* debate. Changes in the expression pattern of genes is often attributed to changes in *cis*-regulatory elements, however, mutations within protein coding regions, often referred to as structural mutations, can also have a dramatic impact on gene expression. Synonymous substitutions in *Drosophila melanogaster* 's alcohol dehydrogenase are known to decrease enzymes leaves by impacting mRNA stability (Carlini and Stephan, 2003).

The genetic code facilitates direct inferences to be made about the phenotypic consequences of nucleotide changes in protein coding regions, such as mutations that result in synonymous or nonsynonymous amino acid changes. More recently, the field of structural genomics has developed algorithms that make 3-dimensional structural predictions of proteins directly from DNA sequence based on similarity to other sequences for which crystallographic or nuclear magnetic resonance structural data are available (Montelione et al., 2009; Nair et al., 2009). However, the identification and phenotypic impact of variation within *cis*-regulatory elements is less tractable than their protein counterparts.

Furthermore, progress in understanding the "genetic toolkit" necessary for normal development has expanded the protein-centric view of molecular evolution to include non-coding regions of the genome that harbor regulatory elements. Briefly, the "genetic toolkit" is a collection of genes that determine the overall animal body plan and the number, identity, and pattern of morphological features (Carroll, 2001). Important caveats have been noted regarding the role of protein coding sequences of developmental genes in the evolution of form and function (Carroll, 2008).

Numerous TFs have been identified that have disparate master regulatory roles during development, such as the role of *sonic hedgehog* in ventral differentiation in the neural tube and later in limb bud formation (Carroll, 2001). Many other master regulatory genes have also been recruited for other roles later in development and therefore, regulate the development of several spatially and temporally disparate structures. A study by Stark et al. (2007) estimated that 67 *Drosophila* transcription factors had, on average, 124 target genes (Carroll, 2008; Stark et al., 2007). The *Drosophila* TF *Twist* has been shown to have approximately 500 target genes that mediate diverse cellular processes including cell proliferation, cell migration, and morphogenesis (Sandmann et al., 2007). More specific to development, *sonic hedgehog* regulates floor plate development in the early vertebrate CNS, and is later recruited to determine digit number and polarity, cerebellum development, and feather bud formation in chickens (McMahon et al., 2003). Therefore, while a newly arisen variant in a TF may provide some evolutionary innovation in one structure during development, that alteration could have dramatic pleiotropic effects in the other pathways that the gene regulates.

While animals have been diversifying for more than 500 million years, the protein components of many signaling systems remain highly similar. For example, 12 *Wnt* genes have been identified in vertebrates, 11 of which are also found in cnidarians (Kusserow et al., 2005). The similarity of orthologs at the sequence level, despite deep species divergence, is continued at the functional level. Functional equivalence of deeply diverged orthologs was demonstrated when *Drosophila Pax-6* knockouts were rescued with mouse-*Pax6* homologs to induce ectopic ommatidia (Halder et al., 1995). Similarly, the cnidarian *Achaete-Scute* homolog

was able to induce sensory organ formation in *Drosophila* and recognize the endogenous protein binding partner *Daughterless* (Grens et al., 1995).

In order for CREs to have a role in evolutionary innovation or constraint they must constitute a major mechanism of gene regulation and present phenotypic consequences when new variants arise. Genes can be regulated by a variety of mechanisms, including chromatin re-modeling, DNA methylation, transcriptional initiation, alternative splicing of RNA, mRNA stability (e.g. 5' capping and 3' polyadenylation), translational controls, covalent post-translational modifications, intracellular trafficking, and protein degradation. Although most genes are regulated at several of these levels, most empirical evidence suggests that transcriptional initiation is the most common and rate-limiting point in regulating gene expression (Carey et al., 2009).

In regards to phenotypic consequences of CRE variation, early screens in *Drosophila* for homeotic mutations demonstrated dramatic changes in abdominal identity through mutations in *cis*-regulatory regions of *Adb-B* (Celniker et al., 1990). Moreover, the range of morphological consequences due to mutations in the promoter region of *Ubx* in *Drosophila* were originally attributed to entirely different genes (Lewis, 1978).

Numerous cases exist where induced mutations have paralleled the phenotypic differences between species. For example, mutations have been induced in *Drosophila* that mimic the identity and number of appendages found in other insects (Carroll, 1995; Raff, 1983), floral anatomy similar to other angiosperms in *Arabidopsis thaliana* (Lawton-Rauth, 2000), and tail anatomy of other nematodes in *Caenorhabditis elegans* (Fitch, 1997). However, as pointed out by Wray et al. (2003), if mutations that alter transcriptional regulation are to be recognized as a major mechanism of evolution then empirical evidence for the genetic

basis of phenotypic differences in natural populations must be demonstrated (Wray et al., 2003).

Definitive cases where changes in transcriptional regulation have driven lineage-specific evolutionary change are limited, but any evolutionary transition germinates from variation within populations. Widespread use of model systems to understand development circumvents many technical obstacles, but can inadvertently project a sense of invariability. Substantial evidence suggests that natural populations from many species harbor significant variation in the spatial pattern of gene expression. In natural populations of rainbow trout, a variant of phosphoglucomutase is expressed in the liver and is associated with increased glycolysis flux in the embryo, developmental buffering, body size advantage, earlier sexual maturity, and overall accelerated prehatching growth (Allendorf et al., 1983; Allendorf et al., 1982). The expression pattern of amylase varies in the midgut of *D. melanogaster* and *D. pseudoobscura*, while the spatial pattern of *Distal-less* varies in correlation with patterns of wing coloration in the butterfly *Bicyclus anynana* (B eldade et al., 2002). In the later case, the expression pattern of *Distal-less* correlated with artificial selection for patterns of wing coloration.

In addition to changes in the spatial pattern of expression, the amount and inducibility of gene expression can be altered at the transcriptional level. Polymorphisms located in *cis* have been associated with changes in the expression level of several gene in natural populations; the aforementioned amylase expression in *D. melanogaster* varies in response to starch diet (Matsuo and Yamazaki, 1984). The level of β-glucuronidase expression varies in natural populations of *Mus domesticus* and is known to differentially respond to endogenous androgen levels (Bush and Paigen, 1992). Changes in prolactin

expression have been documented in the teleost *Oreochromis niloticus* (Streelman and Kocher, 2002), and levels of the cytochrome P450 gene *Cyp6g1* in *D. melanogaster* have been linked to insertion of an upstream transposable element that confers significant insecticide resistance (Daborn et al., 2002). Furthermore, a study of 140 experimentally validated functional *cis*-regulatory polymorphisms for 107 genes in *Homo sapiens* were found to have a two-fold increase in the rate of transcription for >63% of surveyed genes (Rockman and Wray, 2002).

Although intraspecific changes in the spatial pattern, amount, and inducibility of gene expression have been demonstrated, to what extent is this variation heritable? Quantitative genetic surveys based on protein and mRNA expression have been conducted in a variety of species. A strong genetic component was discovered for variation in protein expression, measured electrophoretically, in *Zea mays* (Maize) (Damerval et al., 1994; De Vienne, 2001), *Pinus pinaster* (Maritime Pine) (Costa, 1999), and *Glycine max* (soybean) (Gerber, 2000). Comparison of gene expression, measured as a quantitative trait with microarrays, between natural and lab strains of *Saccharomyces cerevisiae* found that 32% of 570 differentially expressed transcripts mapped to regions within 10 kb of the expressed gene (Brem et al., 2002). A similar study in mice used the correlation between transcript abundance and quantitative trait loci (QTL) to conclude an even higher portion of *cis* located QTLs (Schadt et al., 2003). More broadly, however, statistical techniques have been developed to investigate the regulation of entire genetic networks and their heritability through expression QTL (eQTL) analysis (Kliebenstein, 2009).

Proponents of *cis*-regulatory mediated evolution have propounded that evolution can act more efficiently through changes in *cis* regulatory elements than protein coding regions (Stern, 2000; Wray et al., 2003). Many mutations or alleles at protein coding loci are recessive

and therefore not immediately exposed to selective pressure, particularly if heterosis, the increased fitness of heterozygotes over homozygotes, is at work. However, functional variation in promoter regions is likely to be exposed immediately to selection because of co-dominance; measurements of allele-specific transcript abundance indicate independent regulation (Pastinen et al., 2004; Ronald et al., 2005; Wittkopp et al., 2004). For example, temperature variation in the environment created a latitudinal cline in the proximal promoter region of lactate dehydrogenase-B (*ldh-b*) in populations of *Fundulus heteroclitus* (Mummichog). By comparing functional and non-functional regions of the *ldh-b* promoter, investigators conclusively identified directional selection within promoter elements (Crawford et al., 1999) that were driving population differentiation, specifically, through alteration in Sp1 binding sites (Segal et al., 1999).

Natural variation in expression of the stress-inducible chaperone *Hsp70Ba* gene was ascribed to transcriptional dysregulation induced by multiple transposable element (TE) insertions. Promoter regions polymorphic for two different insertion elements had measurable influence on *hsp70Ba* transcript abundance as revealed by RNase protection assays. Moreover, alleles polymorphic for the TE insertion had phenotypic consequences in inducible thermotolerance and female reproductive success, and therefore immediately subject to purifying selection (Lerman et al., 2003; Walser et al., 2006).

In assessing the evolutionary significance of *cis*-regulatory mutations it remains quite challenging to define casual relationships. Correlational studies between gene expression and phenotypic divergence are valuable for generating hypotheses about the process of evolutionary innovation but transitioning from pattern to process remains dubious (Hoekstra and Coyne, 2007).

Xenopus – A polyploid model system for genetics

The African clawed frogs *Xenopus* and *Silurana*, sister taxa within the subfamily Xenopodinae, family Pipidae (de Sa and Hillis, 1990), are prevalent model organisms (Cannatella and Sa, 1993) and arguably the most well characterized amphibians at the molecular level. However, the distinction between the *Xenopus* and *Silurana* genera are often not fully appreciated in studies that utilize these model systems. The tendency to treat these genera as biological equivalents underscores the continued propagation of a historical artifact despite significant molecular, cytogenetic, and morphological evidence to the contrary. Many experimental studies and major bioinformatic resources (e.g. NCBI) refer to *Xenopus tropicalis* and *Silurana tropicalis* interchangeably. Some of the notable distinctions include, for example, that species along the *Xenopus* lineage have multiples of 18 chromosomes, whereas the *Silurana* lineage includes species with multiples of 20 chromosomes (Cannatella and Trueb, 1988; Evans et al., 2004). Moreover, *Xenopus* and *Silurana* are estimated to have been evolving independently for the last 53-21 million years (Chain and Evans, 2006; Evans et al., 2004), roughly equivalent to the divergence between *Strepsirrhini* (e.g. extant lemurs) and *Haplorrhini* (e.g. lineage leading to humans) primates.

X. laevis have occupied a central role in early vertebrate embryology because of their large embryos, clutch sizes, and easy husbandry in the laboratory. However, more recently *Xenopus* and *Silurana* have gained prominence as an ideal vertebrate model system in evolutionary and developmental biology through extensive EST DNA sequencing, completed sequencing of the *S. tropicalis* genome, advances in transgene efficiency, and prefabricated expression microarrays. While many *Xenopus* species are known to be the

result of bifurcating speciation others are the product of reticulated lineages (i.e. allopolyploidization) (Evans et al., 2004).

The *Xenopus* clade is known to include 10 tetraploid (2N = 36), 5 octoploid (2N = 72), and 2 dodecaploid species that arose through allolopolyploidy (2N = 108) (Schmid and Steinlein, 1991; Tymowska, 1991). Based on the karyotypes of extant *Xenopus* frogs, their evolutionary history is suggested to have originated from an 18 chromosome diploid ancestor. However, WGDs are not unique to clawed frogs, and have occurred in teleost fish (Taylor et al., 2003), salmonid fish (Allendorf, 1984), and the ancestor of all jawed vertebrates (Dehal and Boore, 2005).

Any study that incorporates allele frequency and/or gene expression data in a polyploid model system must exercise interpretive caution. Implementing standard population genetic models based on segregation and independent assortment of alleles may be confounded by whole genome duplication (WGD) events in clawed frogs (Kobel, 1996) that have the potential to produce patterns of polysomic inheritance. Concern stems from the possible formation of multivalents or random bivalents during meiosis that would disrupt the standard diploid model of Mendalian segregation and independent assortment. In autopolyploids, polysomic inheritance is a legitimate concern because each homolog is an exact duplicated copy of its ancestor and unequal chromosome segregation during meiosis is probable. However, allopolyploid genomes arise from the hybridization of two diverged species. Therefore, a disomic pattern of inheritance, consistent with standard diploid population genetic models, may evolve rapidly or immediately following polyploidization. Laboratory generated hybrids of *Xenopus* have suggested disomic inheritance as the prevalent state (Muller, 1977). Moreover, the sequence identity of duplicated paralogs in *X. l. laevis* are

likely to be very diverged as evidenced by the divergence observed between *Xenopus* and *Silurana* over ~50 million years of evolution (Evans et al., 2004). Our analysis is therefore based on a conservative assumption that our loci displays a diploid pattern of inheritance and subsequently amenable to standard population genetic models.

Although cytogenetic evidence suggest disomic inheritance in allopolyploids, the potential exists for homologs from hybridized, yet diverged, species to recombine and obscure the evolutionary history of any given gene or region of the genome. Detecting recombination can be challenging depending on the degree of divergence and extent or rate of heterogeneity of recombination (Posada, 2002). However, investigation at the RAG-1 locus in *Xenopus* and *Silurana* suggest that recombination is infrequent, at lease at this locus, along both lineages (Evans et al., 2005). Similar results were found regardless of the method or parameters specified (e.g. Informative Sites Test, Recombination Detection Program, Geneconv). Therefore, inferences made from the pattern of polymorphism at most loci that assume no recombination may be considered conservative.

Project Overview

We investigated whether the immediate upstream region of *xGAD67* in *Xenopus laevis laevis* has been subject to selective forces. *xGAD67* is a terminal differentiation gene with a tightly regulated spatiotemporal expression pattern during critical periods of neural development, which appears conserved across deeply diverged vertebrates. There is abundant evidence that *cis*-regulatory elements constitute a large source of variation in gene expression in natural populations. The structure and significance of variation in *cis* elements is not well understood, largely due to the difficulty in identifying and determining their

functional effect on overall transcription. However, where there is a conserved biological function, such as the pattern of expression between different species, there is an *a priori* expectation of constraint and marked reduction in variation. We investigated the amount and pattern of nucleotide variation in the 5' flanking region of *xGAD67* in a natural population of *X. l. laevis*[1] individuals using a diverse array of tests for selection.

[1] Most laboratory strains of *X. laevis* are, if fact, inbred lines of the species *Xenopus laevis laevis* from South Africa. Existing phylogenic evidence has identified several *Xenopus laevis* subspecies, including *X. l. victorianus* and *X. l. sudanensis.*

EXPERIMENTAL AND NUMERICAL METHODS

Tissue collection

Genomic DNA samples from 15 *X. laevis laevis* individuals were kindly provided as a gift by B. Evans (McMaster University, Hamilton, Ontario). All samples were collected on the same night in the vicinity of the Lewis Gay Dam, north of Cape Point Nature Reserve in Cape Province, South Africa. Genomic DNA was provided in 1X Tris-EDTA buffer and stored at -20°C. One female and one male from the four closely related outgroups *X. amieti, X. clivii, X. l. sudanensis, and X. muelleri* were obtained commercially (Xenopus Express). Genomic DNA from one individual per outgroup was extracted from ~20 mg of femoral muscle tissue with DNeasy tissue kit (Qiagen) and stored at -80°C.

PCR, cloning, and sequencing

The following primers were used to amplify three regions of the *xGAD67* locus: 5'-ACACCAGCACGTTCTCCATT-3' / 5'-TTGGCAGGGTGTTCTCTTTC-3' (immediate upstream region including a portion of the first exon), 5'-GCCAGAGGTTGTATTTTTCACA-3' / 5'-TGATTTGGAAGCGAAAATCC-3' (intron 15, partial exon 15 and 16), 5'-GCCCCAGTTTTCTTTATGGA-3'/5'-CAGTCCCCAGATTGGTTCTC-3' (intron 17, partial exon 17 and 18). PCR was conducted in a 50 µl reaction volume using 1.25 µl of each [10µM] primer, 0.2 µl of High Fidelity Platinum Taq Polymerase (Invitrogen), 2.5 µl of 10x buffer, 1 µl of 50 mM $MgCl_2$, 1 µl of 10 mM dNTPs, ~225ng of genomic DNA, and brought up to volume with nuclease-free water (Promega). The amplification reaction was performed in a GeneAmp PCR System 9700

(Applied Biosystems) under the following conditions: 1 cycle of 1 min at 94°C; 35 cycles of 20 sec at 94°C for denaturation, 20 sec at 55/53.5/57°C for annealing, and 1-1.5 min at 68°C for extension. PCR products were separated on an ethidium bromide agarose gel and visualized on a FluorChem HD2 analyzer (Alpha Innotech). Amplified products were directly cloned from the PCR reaction mixture into pSC-A-amp/kan vector using the StrataClone PCR Cloning Kit (StrataClone). The ligation reaction consisted of 3 μl of cloning buffer, 2 μl of PCR product, and 1 μl of vector mix (StrataClone), incubated at room temperature for 5 min and immediately transformed or stored at -20°C. Ligation reaction products were transformed into StrataClone SoloPack competent cells according to the manufacturers protocol. Plasmid DNA was recovered from the bacteria with the Wizard Plus SV Miniprep Kit (Promega) and eluted in 30 μl nuclease-free water and stored at 4°C.

DNA samples were prepared for Sanger automated capillary sequencing using BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems). The terminal chain PCR labeling was assembled with ~200 ng template DNA, 3.2 μM of M13 forward and M13 reverse primers, 2 μl of BigDye v3.1 Ready Reaction mix, and brought up to 10 μl with nuclease-free water (Promega). Labeled PCR products were combined with 2.5 μl of 125 mM EDTA, 30 μl of 100% ethanol, and incubated at room temperature for 15 min. Samples were centrifuged for 20 min at 4300 rpm at 4°C in a Mikro 200R. Samples were washed with 200 μl of 70% ethanol and centrifuged an additional 5 min at 4°C. Samples were dried ~5 min on a Savant Speed Vac Plus SC110A and resuspended with 20 μl Hi-Di formamide (Applied Biosystems). Samples were either used immediately for sequencing or stored overnight at 4°C.

DNA samples were loaded onto a 3130 Avant Genetic Analyzer equipped with a 3130 Capillary Array 50 cm column and high performance polymer Pop7 (Applied Biosystems). Internal sequencing primers were designed as necessary in order to ensure complete bi-directional sequencing. Sequence fragments were assembled with Sequencer 4.9 (Gene Codes) using default parameters. Assemblies were exported from Sequencer in FASTA format and stored for general viewing and organization on Vector NTI (Invitrogen) software. The identity of each allele was confirmed using sequence data from multiple clones (2-8) aligned with Clustal running locally in Vector NTI.

Population and outgroup sequences for introns 15 and 17 were aligned using ClustalX 2.0.11 (http://www.clustal.org/) under default settings (Larkin et al., 2007) and manually adjusted using Mesquite (http://mesquiteproject.org) in order to minimize the number of unnecessary gaps and maximize local regions of identity (Maddison and Maddison, 2009). The upstream region of *xGAD67* from our *X. l. laevis* population and outgroup samples was aligned with Threaded Blockset Aligner (TBA) using default parameters (Blanchette et al., 2004) with a binary tree of our five *Xenopus* species described in a modified Newick format (details in Appendix) and manually adjusted in Mesquite. Multiple sequence alignments for each locus were generated in PHYLP and FASTA formats for downstream applications.

Sequence Statistics

Software kindly provided by R. Haygood (Duke University, Durham, NC) was used to generate our summary statistics and selection results. R.H. programs were executed in MacPython 2.5 within the Biopython framework (http://www.biopython.org; build 2.5.2); source code is provided in the Appendix.

Selection Analysis

The conserved expression pattern of *GAD67* across distantly related vertebrates suggests that purifying selection should be acting in noncoding regions that harbor regulatory elements. We have therefore employed several test statistics of selection in order to determine if the immediate 5' upstream region of *xGAD67* has experienced selective pressure. All of our models assume an ideal Wright-Fisher population (Fisher, 1930; Wright, 1931) and are equitably applied to our allopolyploid system for reasons mentioned earlier. Thus, populations have discrete and non-overlapping generations, with $2N$ number of genes and N number of diploid individuals, constant population size where all individuals have equal fitness, no geographic or social structure, and no recombination except where indicated (Hein et al., 2005).

The signature of directional selection, either positive or negative, can be inferred as a significant deviation from a neutral expectation based on predicted patterns of polymorphism and substitution. Two test statistics, Tajima's D and Fay and Wu's H, attempt to detect selection by using two different estimators of the scale mutation rate (θ), also referred to as a population summary statistics of variation. Watterson introduced the first measure of a general class of θ estimators where $\theta = 4N_e\mu$; N_e is the effective population

size and μ is the neutral mutation rate of the population per generation, as an estimate of genetic diversity based on a scaled mutation rate (Watterson, 1975). For a population at equilibrium, Watterson derived the expectation for the number of segregating sites (S)

$$E(S) = a_1\theta_S,$$

and variance

$$V(S) = a_1\theta_S + a_2\theta^2$$

where

$$a_1 = \sum_{S=1}^{n-1} \frac{1}{S}$$

$$a_2 = \sum_{S=1}^{n-1} \frac{1}{S^2}$$

of θ_S based on the number of segregating sites (S) in a Wright-Fisher population model of n randomly selected haploid individuals (also referred to as gametes or haplotypes). When the sample size $n << N_e$ then S becomes the sum of n - 1 independent, geometrically distributed random variables (Watterson, 1975). An infinite sites model of mutation is assumed where each locus contains a large number of independent sites, mutations are rare events, and back mutations do not occur. In order to detect the action of selection in the *xGAD67* promoter, we have obtained sequence data from a population of *X. l. laevis* individuals and one representative sequence from each closely related outgroup species: *X. amieti, X. clivii, X. l. sudanensis, and X. muelleri.* We used software written by R.H. to compute Tajima's *D*, Fu and Li's *D*, and Fay and Wu's *H*, and Hudson, Kreitman, Aguade (HKA) test statistics. These summary statistics with the exception of HKA utilize the difference between various estimators of θ and are detailed below. Our analyses assume an ideal Wright-Fisher population at stationarity.

We estimated Tajima's *D* test statistic at the *xGAD67* locus with 26 *X. l. laevis* samples (Tajima, 1989). Tajima proposed the θ_π estimator based on the average number of pairwise differences, π, expected in a random sample drawn from a population (Tajima, 1983) and determined the expected number of pairwise differences

$$E(\pi) = \theta_\pi = \frac{\sum\sum \pi_{ij}}{\binom{n}{2}},$$

and variance,

$$V(\pi) = b_1\theta_\pi + b_2\theta_\pi^2,$$

where

$$b_1 = \frac{n+1}{3(n-1)}$$

$$b_2 = \frac{2(n^2+n+3)}{9n(n-1)}$$

for a neutrally evolving locus. Therefore, θ_S and θ_π are estimators of low frequency (i.e. rare variants) and intermediate frequency (i.e. common variants) alleles, respectively. The biological significance of these two different estimators of θ is their relationship in the presence of selection. Under neutrality there should be no difference between θ_S and θ_π ($0 = \theta_\pi - \theta_S$), however purifying selection will act to keep new variants at low frequency whereby generating an excess of extreme low frequency alleles in a population $\theta_S > \theta_\pi$. The presence of positive selection, conversely, drives new advantageous alleles to higher frequency and $\theta_\pi > \theta_S$. Tajima proposed the test statistic

$$D = \frac{\theta_\pi - \frac{\theta_S}{a_1}}{\sqrt{\text{var}\left(\theta_\pi - \frac{\theta_S}{a_1}\right)}},$$

which has a β-distribution under neutrality. While Tajima's *D* can be an informative statistic for interpreting polymorphism data there are several limitations. Demographic perturbations, such as population bottlenecks, recent expansions, hitchhiking, and mutational rate heterogeneity can produce a pattern of polymorphism similar to selection. For example, population bottlenecks will greatly increase θ_S while reducing $\theta_\pi$and result in negative *D* values that mimic purifying selection. The difficulty in distinguishing demographic influences from selection can be partly mitigated by using several different measures of selection and investigating multiple loci. For example, a dramatic drop in population size with display an excess of low frequency alleles across all loci while an excess of low frequency allele generated by a strong selective sweep will only effect tightly link loci (i.e. genetic hitchhiking). The statistical significance of our results was determined by coalescent simulation detailed in a subsequent section (*see* Coalescent Simulation).

We computed Fu and Li's *D* test statistic of selection on our population data with each outgroup species (Fu and Li, 1993). Fu and Li proposed to examine the distribution of mutations, also referred to as the site-frequency spectrum, within a genealogical framework. Similar to Tajima's *D*, Fu and Li posited that deviations from neutrality can be detected by the differences between the number of mutations expected in internal (η_i) and external (η_e) branches of a genealogy, and derived the relationship of these expectances to θ estimators. The distribution of mutations along a genealogy can be categorized as internal if they occur

in the older regions (i.e. outgroup branches) of the tree, or external if they occur in newer parts of the tree (ingroup population branches). θ estimates of internal, i, and external, e, mutations are

$$\theta_i = \frac{\eta_i}{(a_n - 1)},$$
$$\theta_e = \eta_e$$

respectively, where

$$a_n = \sum_{k=1}^{n-1} \frac{1}{k}.$$

In a sample of n sequences with i = 1, 2, ... m nucleotide sites. The total numbers of mutations in the sample and the number of mutations in the external branches of the genealogy are

$$\eta = \sum_{i}^{m} s_i \text{ and } \eta_e = \sum_{i}^{m} e_i,$$

respectively; s_i are the number of segregating sites in the population-outgroup alignment minus one, e_i are the number of singleton sites between the ingroup and outgroup sequences (ingroup segregating sites ignored). The null hypothesis, under the neutral model, states that there should be no difference between θ_i and θ_e estimates, therefore the normalized test statistic becomes

$$D = \frac{\eta - a_n \eta_e}{\sqrt{u_D \eta + v_D \eta^2}}$$

where

$$v_D = 1 + \frac{a_n^2}{b_n + a_n^2}\left(c_n - \frac{n+1}{n-1}\right)$$
$$u_D = a_n - 1 - v_D$$

and

$$b_n = \sum_{k=1}^{n-1} \frac{1}{k^2}.$$

Consider a sample of sequences from a population and an outgroup organized into a genealogy, where internal branches represent mutations in "older" regions and external branches consist of "new" mutations. Under this model, the action of adaptive or purifying selection would generate an excess of mutations in the external branches of the genealogy. We conducted the test with n = 26 samples from the *X. l. laevis* population, which served as external branches. Related species *X. amieti, X. clivii, X. l. sudanensis, and X. muelleri* were treated as n_e outgroups to define internal older branches of the genealogy. The statistical significance of our results was determined by coalescent simulation detailed in a subsequent section (*see* Coalescent Simulation).

We also estimated Fay and Wu's H statistic as the third independent survey of the site-frequency spectrum (Fay and Wu, 2000) at the *xGAD67* locus. Analogous to the preceding tests, the H statistic is based on the difference between two estimators of θ, specifically θ_π and θ_H (Fu, 1995), where

$$\theta_H = \sum_{i=1}^{n-1} \frac{2S_i i^2}{n(n-1)},$$

and

$$\theta_\pi = \sum_{i=1}^{n-1} \frac{2S_i i(n-i)}{n(n-1)}$$

where S_i is the number of segregating sites found in a sample size of n with a frequency i times. Noteworthy is that as i approaches $n/2$, alleles of intermediate frequency preferentially influence θ_π, while as i approaches n (i.e. numerous derived alleles at high frequency) the

effect has a greater contribution on θ_H. Originally proposed to overcome the limitations of Tajima's D in detecting hitchhiking events, the θ_H estimator is influenced by an excess of high-frequency alleles characteristic of strong selection. Fay and Wu's H statistic is the difference between θ_π and θ_H, divided by the variance of these estimators, and expected to be zero at a neutrally evolving locus. Negative values of H indicate that there is an excess of high frequency alleles, the signature that tightly linked low frequency alleles have been driven to high frequency by genetic hitchhiking. As stated previously, derived and ancestral sequences were defined by an ingroup population of *X. l. laevis* individuals and several closely related outgroups (e.g. *X. amieti, X. clivii, X. l. sudanensis, and X. muelleri*).

Finally, we used software by R.H. to conduct the Hudson, Kreitman, and Aguade (HKA) test on the 5' upstream region of *xGAD67* using intron 17 as a neutral proxy (Hudson et al., 1987). The HKA test of a neutral model predicts that intraspecies polymorphism and interspecies divergence are the products of the neutral mutation rate. A Fisher-Wright ideal population and infinite-sites model of mutation are assumed. For each locus i (i = 2) the number of mutations per sample per generation is Poisson distributed. There is no recombination within loci, but free recombination is allowed to occur between loci. The ingroup and outgroup populations are assumed at equilibrium with population sizes $2N$ and $2Nf$, respectively. The population and outgroup species diverged T generations in the past from a single ancestral population. Our analysis included 24 *X. l. laevis* sequences representing species A, also referred to as the ingroup, while a single sequence from species B will represent the outgroup species as done in Hudson *el al.* (1987) example data set. Several outgroup species will be considered independently. We examined i = 2 loci, the

promoter region of *xGAD67* and intron 15, for n = 24 random samples. Hudson *et al.* (1987) derived the expectation and variance at a neutral locus for a population of samples as

$$E(S_i^A) \approx \theta_i C(n_A),$$

$$Var(S_i^A) \approx E(S_i^A) + \theta^2 \sum_{j=1}^{n_a-1} \frac{1}{j^2},$$

and for an outgroup as

$$E(D_i) = \theta_i \left(T + \frac{(1+f)}{2} \right),$$

$$Var(D_i) = E(D_i) + \left(\theta_i \frac{(1+f)}{2} \right)^2,$$

where

$$C(n) = \sum_{j=1}^{n-1} \frac{1}{j};$$

S_i^A is the number of nucleotide sites in species A (i.e. *X. l. laevis*) that are polymorphic at loci i, D_i^A is the number of substitution between the ingroup population and an outgroup species, where each outgroup (*X. amieti, X. clivii, X. l. sudanensis, and X. muelleri*) is examined separately, $\theta_i = 4N\mu_i$, $T = T' / 2N$. The test statistic

$$\chi^2 = \sum_{i=1}^{L} \frac{\left(S_i^A - E(S_i^A)^2\right)}{Var(S_i^A)} + \sum_{i=1}^{L} \frac{\left(D_i - E(D_i)^2\right)}{Var(D_i)},$$

was proposed as a goodness-of-fit of the empirical results to the neutral model.

Estimating ρ

Assumptions about the rate of recombination can have important implications for interpreting polymorphism. The distribution of test statistics extracted from simulated data sets can be highly influenced by linkage. Likelihood methods have been developed to analyze samples with multiple linked polymorphic sites (Kuhner et al., 2000; Nielsen, 2000). We used Hudson's maxhap program (http://home.uchicago.edu/~rhudson1) to determine two-site configurations for a sample of *X. l. laevis* sequences with an ancestral state specified by several closely related species. Our approach is to compare empirically determined two-site sample configurations with two-site sample distributions under a neutral model of evolution (Hudson, 2001). The objective was to estimate ρ, where $\rho = 2\, r\, N_e$; where N_e is the effective population size and r is the rate of recombination (Hill, 1975). The estimator ρ, which is a scaled population estimate of recombination, was calculated at each locus for each outgroup species (see Appendix for details). While the most conservative estimate of recombination is $\rho = 0$, it may cause the neutral or null hypothesis of a given test statistic not to be rejected when, in fact, selection may be present. The program exhap was used to convert a set of haplotype data into an input file with the appropriately formatted pairs data. The program maxhap then estimates ρ from our phased haplotype polymorphism data.

Coalescent Simulation

Our ability to distinguish variation generated by selection and variation generated by random processes was assessed by coalescent simulation of a stable, neutrally evolving panmictic population (Hudson, 2002). Source code and documentation are available online (http://home.uchicago.edu/~rhudson1) while the details of our command-line instructions executed in a UNIX shell for each locus are provided in the Appendix. For the upstream

region and intron 17, we conducted 100,000 independent, replicate coalescent simulations for a Wright-Fisher ideal population of N haploid individuals at a locus of specified length and fixed number of segregating sites (Depaulis et al., 2001; Rosenberg and Nordborg, 2002; Wall and Hudson, 2001). Standard small sample approximations of the coalescent are assumed, where the sample size $n << N_e$. An infinite sites model of mutation is assumed where multiple hits and back mutations do not occur.

A random genealogical history is generated for each replicate sample at zero and estimated rates of recombination. Mutations, which are parameterized as $4 N_e \mu$, where N_e is the effective diploid population size, and μ is the neutral mutation rate, are Poisson distributed along a genealogy with mean equal to the product of the mutation rate and branch length. For each independent replicate simulation π, Tajima's D, Fay and Wu's H, and Fu and Li's D values were extracted with Hudson' s sample_stats program. The frequency distributions of these test statistics were plotted as histograms using SPSS Statistics 12.0 and compared to our empirically derived results.

RESULTS

Alignment of the xGAD67 upstream region, intron 15, and intron 17

The amount and significance of sequence variation in noncoding, potentially regulatory, DNA was investigated in the upstream region of *xGAD67* in *Xenopus*. Approximately 1.2 kbps upstream of the *xGAD67* initiating methionine codon was cloned and sequenced for 27 alleles derived from 30 individuals of a natural population of South

African *X. l. laevis* clawed frogs. Three alleles were excluded from our analysis based on poor quality sequence data or incomplete contigs assemblies. These alleles will be included once the data become available. Additionally, one allele was obtained from several closely related *Xenopus* species, including *X. l. sudanensis, X. amieti, X. clivii, and X. muelleri* at the homologous *xGAD67* locus. An alignment of all 31 sequences was originally attempted with ClustalX 2.0.11 under default parameters (*see* Appendix). The resulting alignment presented with numerous gaps of variable length. The pattern and variable length of gaps suggested that many could be spurious, resulting from the inability of ClustalX to aligned closely related sequences with potentially large indels (insertion or deletion). A poor alignment could artificially inflate the number and length of indel polymorphisms or obscure informative segregating sites. Attempts were made to find a more parsimonious alignment that would optimize local regions of identity and minimize the number of reported gaps. Various combinations of gap opening and extension penalties were explored, in addition to manual adjustments in Mesquite, but all failed to give an alignment with an appreciable reduction in gaps.

The Clustal-class aligners, such as T-Coffee (Notredame et al., 2000), MAFFT (Katoh et al., 2002), and MUSCLE (Edgar, 2004) implement global, progressive-programming algorithms, and were originally designed to analyze protein sequences. These global aligners may be inappropriate for aligning non-coding regions where the rates of nucleotide and indel length polymorphisms are higher than in protein coding regions (Haddrill et al., 2008; Halligan et al., 2004; Liang et al., 2008). Therefore, a dynamic-programming alignment was attempted on our sample set with Threaded Blockset Aligner (TBA), originally designed for genome assembly, using default parameters (*see* Appendix.).

Inspection of the TBA alignment suggests that a more parsimonious alignment was obtained. The TBA output alignment was manually adjusted with Mesquite and used for downstream analysis (Fig 1.). Inspection of the alignment demonstrates, as expected, that one large indel event was obscured in the previous two alignments by the introduction of numerous unnecessary, and biologically meaningless, gaps. An alignment of 1,308 bps was obtained for the upstream region of *xGAD67* with 973 informative sites used to compute the several summary statistics and our selection analysis, which regard single nucleotide polymorphisms as informative. Detailed in a following subsection, we attempted to recapture the potential significance of indels with several measures of simple indel diversity and selection, including Tajima's *D* and Fu and Li's *D* test statistics modified for simple indels.

```
                ********** ********************************** ************************
XLL1        ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL2        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL3        --ACCAGCACGTTCTCCATCCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL4        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL5        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL6        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL7        ----CAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL8        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL9        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL10       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL11       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL12       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL13       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL14       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL15       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL16       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL17       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL18       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL19       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL20       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL21       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL22       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL23       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL24       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL25       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL26       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL27       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XA          --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XC          --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLS         --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XM          --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC

                ******* ********************************** ********* ****************************
XLL1        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL2        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL3        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL4        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL5        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL6        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL7        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL8        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL9        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL10       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL11       AAAGCACGAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL12       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL13       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL14       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL15       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL16       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL17       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL18       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL19       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL20       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL21       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL22       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL23       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL24       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL25       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL26       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL27       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XA          AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XC          AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCCCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLS         AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XM          AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTGGAAATAGCACTAATGTCGGTATAAACAT
```

```
          ********************* ** ****** * ********** ******  ****** ************** ********
XLL1      TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  240
XLL2      TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL3      TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL4      TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL5      TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL6      TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL7      TAATGGAGAAGGATTTTGTAAGAGAAAGGAATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  236
XLL8      TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL9      TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL10     TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL11     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL12     TAATGGAGAAGGATTTTGTAAGAAAAAGGGA-TTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTCCGAAGCTT  237
XLL13     TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL14     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTTTTCGCAGTCACATTTCGAAGCTT  238
XLL15     TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGAACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT  236
XLL16     TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGAACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT  236
XLL17     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL18     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL19     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL20     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL21     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL22     TAATGGAGAAGGATTTTGTAAGAAAAAGGGACTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL23     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL24     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL25     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL26     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLL27     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XA        TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGGACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT  236
XC        TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XLS       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT  238
XM        TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGGACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT  236

          ********************** * ******************** ******** ******* ** ***************
XLL1      GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  317
XLL2      GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA  318
XLL3      GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL4      GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL5      GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA  318
XLL6      GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTGGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL7      GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  313
XLL8      GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA  318
XLL9      GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL10     GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL11     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA  318
XLL12     GAAGGTCAAAGTGTCTCACTAGAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA  317
XLL13     GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL14     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA  318
XLL15     GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  313
XLL16     GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  313
XLL17     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL18     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL19     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL20     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL21     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL22     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL23     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL24     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL25     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCTTGCTGCTCCTATAAATT---  315
XLL26     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLL27     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XA        GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  313
XC        GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  315
XLS       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGTTGCTCCTATAAATTTCA  318
XM        GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---  313
```

```
          * ****************************** ************** *** ************* *** ***** *****
XLL1      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   395
XLL2      CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   396
XLL3      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL4      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL5      CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   396
XLL6      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL7      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   391
XLL8      CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   396
XLL9      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL10     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL11     CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   396
XLL12     CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   395
XLL13     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL14     CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTTTACCTATGTG   397
XLL15     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCCCTAAAACAAAAA-TAC-TTACCTATGTG   391
XLL16     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   391
XLL17     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL18     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL19     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL20     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL21     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL22     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCCATGTG   393
XLL23     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL24     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL25     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTGCCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL26     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLL27     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XA        CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   391
XC        CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   393
XLS       CCATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG   396
XM        CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAAATAC-TTACCTATGTG   392

           ************************  *********** ****** ** ***************** **************
XLL1      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   474
XLL2      CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC   475
XLL3      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAGGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL4      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL5      CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC   475
XLL6      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL7      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   470
XLL8      CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC   475
XLL9      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL10     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   473
XLL11     CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC   475
XLL12     CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC   474
XLL13     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL14     CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC   476
XLL15     CTGGCATTTTGGACATTTGTAGACACGACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   470
XLL16     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   470
XLL17     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL18     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL19     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL20     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL21     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL22     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL23     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL24     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL25     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL26     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XLL27     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   472
XA        CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   470
XC        CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGCCTGTGTATACTGTC   472
XLS       CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC   475
XM        TTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC   471
```

```
          ** * ** ** ********* ***************** ********** ** ***   **********************
XLL1      CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  554
XLL2      CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTGGACTGAGAATTAAAATAGGCCC  555
XLL3      CCTTCCCCGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL4      CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL5      CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTGGACTGAGAATTAAAATAGGCCC  555
XLL6      CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL7      CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  550
XLL8      CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTGGACTGAGAATTAAAATAGGCCC  555
XLL9      CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCGGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL10     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  553
XLL11     CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTAGACTGAGAATTAAAATAGGCCC  555
XLL12     CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTGGACTGAGAATTAAAATAGGCCC  554
XLL13     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL14     CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTGGACTGAGAATTAAAATAGGCCC  556
XLL15     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCTGGACTGAGAATTAAAATAGGCCC  550
XLL16     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCTGGACTGAGAATTAAAATAGGCCC  550
XLL17     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL18     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL19     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL20     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL21     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL22     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL23     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL24     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL25     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL26     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLL27     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XA        CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCTGGACTGAGAATTAAAATAGGCCC  550
XC        CCCTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCCGGACTGAGAATTAAAATAGGCCC  552
XLS       CCTTTCCTGAGCATTAGCAATAGGAGATTTAGTGAAATAGGTCCTGGTATCAGGCTCGGACTGAGAATTAAAATAGGCCC  555
XM        CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCTGGACTGAGAATTAAAATAGGCCC  551

          ********** *********
XLL1      TGGCATTTCAGGTACACAGAGACCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCC  634
XLL2      TGGCATTTCAGGTACACAGAGGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCC  635
XLL3      TGGCATTTCAGGTACACAGAGACCCAAACAGCCCCACCTGCCCACTAGATACTGACTTTCTATGGCACCTTATAGCAGCC  632
XLL4      TGGCATTTCAGGTACACAGAGACCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCC  632
XLL5      TGGCATTTCAGGTACACAGAGGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCC  635
XLL6      TGGCATTTCAGGTACACAGAGACCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCC  632
XLL7      TGGCATTTCAGGTACACAGAGACCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCC  630
XLL8      TGGCATTTCAGGTACACAGAGGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCC  635
XLL9      TGGCATTTCAGGTACACAGAGACCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCC  632
XLL10     TGGCATTTCAGGTACACAGAGACCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCC  633
XLL11     TGGCATTTCAGGTACACAGAGGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCC  635
XLL12     TGGCATTTCAGGTACACAGAG----------------------------------------------------------  575
XLL13     TGGCATTTCAGGTACACAGAG----------------------------------------------------------  573
XLL14     TGGCATTTCAGGTACACAGAG----------------------------------------------------------  577
XLL15     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  580
XLL16     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  580
XLL17     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XLL18     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XLL19     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XLL20     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XLL21     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XLL22     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XLL23     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XLL24     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XLL25     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XLL26     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XLL27     TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XA        TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  580
XC        TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  582
XLS       TGGCATTTCAGTTACACAGAG----------------------------------------------------------  576
XM        TGGCATTTCAGGTACACAGAGACCCAAACA-------------------------------------------------  581
```

```
XLL1        CCTCTGGCATTTGCCAGAACCCAC----------------------------------------------------------
XLL2        CCTCTGCCATTTGCCAGAACCCAC----------------------------------------------------------
XLL3        CCTCTGGCATTTGCCAGAACCCAC----------------------------------------------------------
XLL4        CCTCTGGCATTTGCCAGAACCCAC----------------------------------------------------------
XLL5        CCTCTGCCATTTGCCAGAACCCAC----------------------------------------------------------
XLL6        CCTCTGGCATTTGCCAGAACCCAC----------------------------------------------------------
XLL7        CCTCTGGCATTTGCCAGAACCCAC----------------------------------------------------------
XLL8        CCTCTGCCATTTGCCAGAACCCAC----------------------------------------------------------
XLL9        CCTCTGGCATTTGCCAGAACCCAC----------------------------------------------------------
XLL10       CCTCTGGCATTTGCCAGAACCCAC----------------------------------------------------------
XLL11       CCTCTGCCATTTGCCAGAACCCAC----------------------------------------------------------
XLL12       ----------------------------------------------------------------------------------
XLL13       ----------------------------------------------------------------------------------
XLL14       ----------------------------------------------------------------------------------
XLL15       ------------------------CAGGTCTGGGCTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCGCA
XLL16       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCGCA
XLL17       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XLL18       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XLL19       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XLL20       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XLL21       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XLL22       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XLL23       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XLL24       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XLL25       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XLL26       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XLL27       ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XA          ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCGCA
XC          ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCA
XLS         ----------------------------------------------------------------------------------
XM          ------------------------CAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCGCA


XLL1        ----------------------------------------------------------------------------------
XLL2        ----------------------------------------------------------------------------------
XLL3        ----------------------------------------------------------------------------------
XLL4        ----------------------------------------------------------------------------------
XLL5        ----------------------------------------------------------------------------------
XLL6        ----------------------------------------------------------------------------------
XLL7        ----------------------------------------------------------------------------------
XLL8        ----------------------------------------------------------------------------------
XLL9        ----------------------------------------------------------------------------------
XLL10       ----------------------------------------------------------------------------------
XLL11       ----------------------------------------------------------------------------------
XLL12       ----------------------------------------------------------------------------------
XLL13       ----------------------------------------------------------------------------------
XLL14       ----------------------------------------------------------------------------------
XLL15       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL16       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL17       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL18       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL19       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL20       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL21       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL22       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL23       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL24       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL25       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL26       ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLL27       ATCAGCCCCCACCAGCCCACTAAATACCGACCTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XA          ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XC          ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
XLS         ----------------------------------------------------------------------------------
XM          ATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACA
```

```
XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    ----------------------------------------------------------------------------------------
XLL5    ----------------------------------------------------------------------------------------
XLL6    ----------------------------------------------------------------------------------------
XLL7    ----------------------------------------------------------------------------------------
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   ----------------------------------------------------------------------------------------
XLL12   ------------------GCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL13   ------------------ACCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL14   ------------------GCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL15   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL16   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL17   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL18   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL19   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL20   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL21   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL22   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL23   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL24   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL25   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL26   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLL27   GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XA      GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XC      GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XLS     ------------------GCCCAAACAGCCCCACCAGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT
XM      GATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCT

                                  ************   * *************** ******************** ******
XLL1    --------------------AGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL2    --------------------AGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL3    --------------------AGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL4    --------------------AGATTGCCAGTCTGGGTCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL5    --------------------AGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL6    --------------------AGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL7    --------------------AGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL8    --------------------AGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL9    --------------------AGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL10   --------------------AGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL11   --------------------AGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL12   CTGCCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL13   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL14   CTGCCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XLL15   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAACCATACATCCCTTCTCTATCTGTTTGCT
XLL16   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLL17   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLL18   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLL19   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLL20   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLL21   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLL22   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLL23   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLL24   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLL25   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLL26   CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLL27   CTGGCATTTGCCAGAATCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XA      CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XC      CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
XLS     CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCAAGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTATTTGCT
XM      CTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCT
```

```
          **** ************************* ****************************************** *******
XLL1      CAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL2      CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL3      CAGTTAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL4      CAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL5      CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL6      CAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL7      CAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL8      CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL9      CAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL10     CAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL11     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL12     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL13     CAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL14     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL15     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL16     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL17     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL18     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL19     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL20     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL21     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL22     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL23     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL24     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL25     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL26     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLL27     CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XA        CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XC        CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA
XLS       CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCATTTGATTA
XM        CAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTA

          ************************ ************************************ ******* **********
XLL1      GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL2      GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL3      GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL4      GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL5      GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL6      GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL7      GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL8      GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL9      GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL10     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAC
XLL11     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL12     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL13     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACGTTAGGTTTTTAACTTGCAT
XLL14     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL15     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL16     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL17     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL18     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL19     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL20     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL21     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL22     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL23     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL24     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL25     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL26     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLL27     GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XA        GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XC        GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
XLS       GTGGAGTGGAATCTGTGTGATCAAGTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTCTTAACTTGCAT
XM        GTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCAT
```

```
          ****************** ************************************ * ************ ** *******
XLL1      GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL2      GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCCTGTCATATAA
XLL3      GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL4      GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL5      GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL6      GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCTCAAAGGCAGGCCATGTCATATAA
XLL7      GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL8      GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCCTGTCATATAA
XLL9      GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL10     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL11     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCCTGTCATATAA
XLL12     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCCTGTCATATAA
XLL13     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL14     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCCTGTCATATAA
XLL15     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL16     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL17     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL18     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL19     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL20     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL21     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL22     GGCTGCTTTCCATTACCCGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACCCCCAAAGGCAGGCCATGTCATATAA
XLL23     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL24     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL25     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL26     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLL27     GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XA        GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XC        GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA
XLS       GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGCCATATAA
XM        GGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAA

          ********************** **********  ****** ******** *********** ******************
XLL1      ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL2      ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL3      ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL4      ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL5      ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL6      ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGGGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL7      ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL8      ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL9      ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGCATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL10     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL11     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCCATTGCTTGATAGCCTGAC
XLL12     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL13     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL14     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL15     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL16     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL17     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL18     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL19     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL20     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL21     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL22     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL23     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL24     ATAGTCTTTCCACTGGCCTCACTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL25     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL26     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLL27     ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XA        ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XC        ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XLS       ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
XM        ATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGAC
                                  +1
```

```
                ***************************
XLL1            CATGCTCAGGAAAGAGAACACCCTGCCA
XLL2            CATGCTCAGGAAAGAGAACACCCTGCC-
XLL3            CATGCTCAGGAAAGAGAACACCCTGCC-
XLL4            CATGCTCAGGAAAGAGAACACCCTGCC-
XLL5            CATGCTCAGGAAAGAGAACACCCTGC--
XLL6            CATGCTCAGGAAAGAGAACACCCTGCC-
XLL7            CATGCTCAGGAAAGAGAACACCCTGCC-
XLL8            CATGCTCAGGAAAGAGAACACCCTGCC-
XLL9            CATGCTCAGGAAAGAGAACACCCTGCC-
XLL10           CATGCTCAGGAAAGAGAACACCCTGC--
XLL11           CATGCTCAGGAAAGAGAACACCCTGC--
XLL12           CATGCTCAGGAAAGAGAACACCCTGCC-
XLL13           CATGCTCAGGAAAGAGAACACCCTGCC-
XLL14           CATGCTCAGGAAAGAGAACACCCTGCC-
XLL15           CATGCTCAGGAAAGAGAACACCCTGCC-
XLL16           CATGCTCAGGAAAGAGAACACCCTGCC-
XLL17           CATGCTCAGGAAAGAGAACACCCTGCC-
XLL18           CATGCTCAGGAAAGAGAACACCCTGCC-
XLL19           CATGCTCAGGAAAGAGAACACCCTGC--
XLL20           CATGCTCAGGAAAGAGAACACCCTGC--
XLL21           CATGCTCAGGAAAGAGAACACCCTGC--
XLL22           CATGCTCAGGAAAGAGAACACCCTGC--
XLL23           CATGCTCAGGAAAGAGAACACCCTGC--
XLL24           CATGCTCAGGAAAGAGAACACCCTGC--
XLL25           CATGCTCAGGAAAGAGAACACCCTGCC-
XLL26           CATGCTCAGGAAAGAGAACACCCTGCC-
XLL27           CATGCTCAGGAAAGAGAACACCCTGCC-
XA              CATGCTCAGGAAAGAGAACACCCTGCC-
XC              CATGCTCAGGAAAGAGAACACCCTGCC-
XLS             CATGCTCAGGAAAGAGAACACCCTGCC-
XM              CATGCTCAGGAAAGAGAACACCCTGCC-
```

Figure 1. Manually adjusted TBA alignment of the upstream region of *xGAD67*. Manual adjustments were performed in Mesquite in order to minimize the number of unnecessary gaps. Local regions of conservation not aligned by TBA were identified and adjusted as necessary, often eliminating gaps. The transcriptional start site is indicated by +1 and the initiating methionine is underlined. Sequences proceed from 5' to 3' and include a portion of the first exon. Conserved sites are indicated by an asterisk (*).

Introns 15 and 17 of *xGAD67* were amplified and sequenced for 24 and 27 alleles, respectively, from our population sample. There were no PCR products recovered for intron 15 and 17 from several individuals. The intronic alleles were aligned with ClustalX 2.0.11 using default parameters and manually adjusted with Mesquite to minimize the number of gaps, as described earlier for the upstream region (Fig 2 and Fig 3). Evidence from mammals suggests that introns are amongst the least constrained regions in the genome and therefore provide the best empirical measure of neutrality (2005; Hellmann et al., 2003; Keightley et al., 2005). The ClustalX alignment of intron 15 was slightly different with the inclusion of each *Xenopus* outgroup, therefore the results for each outgroup are reported (Table 2.). Intron 15 was determined to be 1,888 bps in length, however the number of informative sites ranged from 1,190 to 1,322 bps due to indel presence/absence in the outgroup. Alignment of intron 17 was uniform across all four *Xenopus* species and yielded 692 bps in total length with 654 considered as informative for analyzing the site-frequency spectrum.

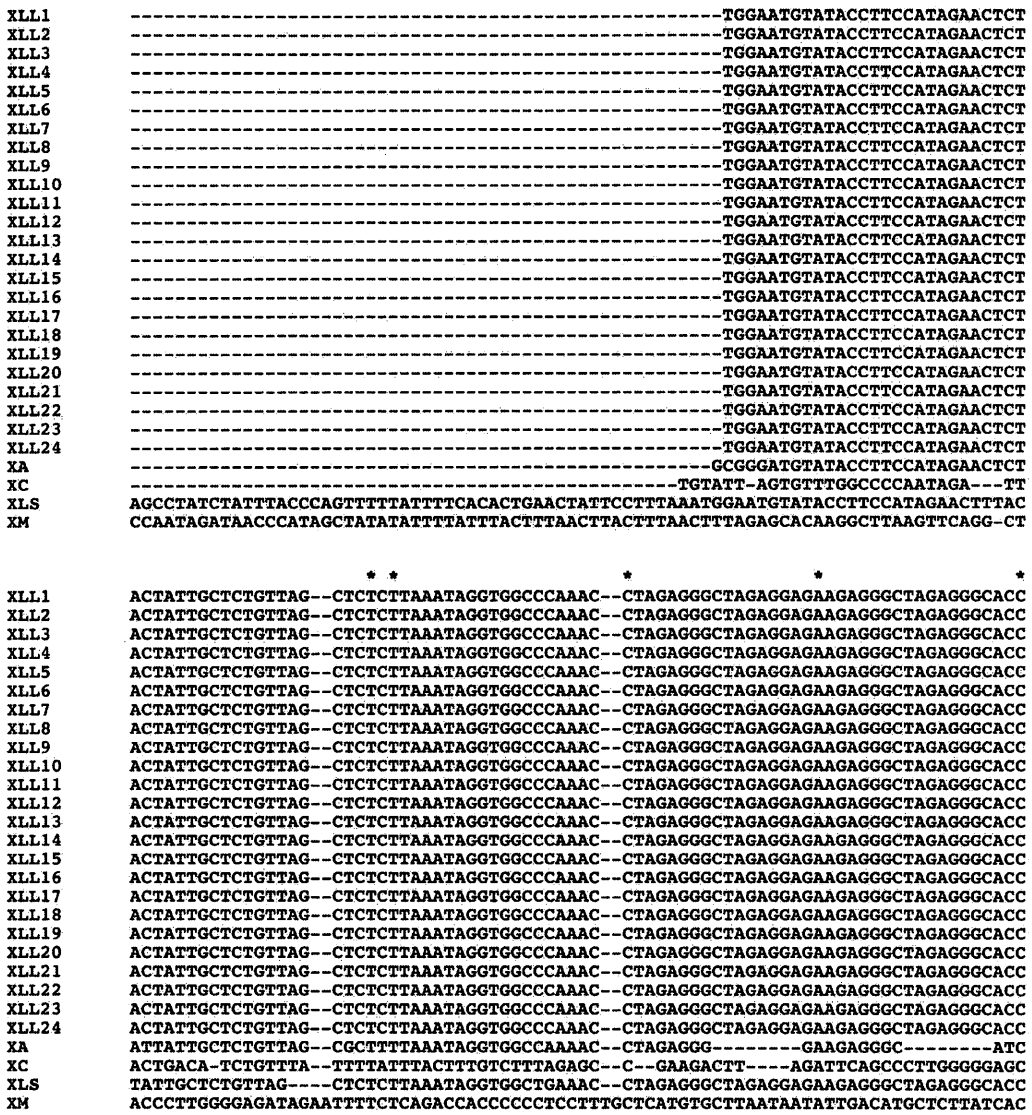
```
                ****************************************** *** ***** ****  *********  ****
XLL1            TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL2            TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL3            TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL4            TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL5            TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL6            TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL7            TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL8            TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL9            TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL10           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL11           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL12           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL13           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL14           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL15           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL16           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL17           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL18           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL19           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL20           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL21           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL22           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL23           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XLL24           TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XA              TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTAGTAGCAGGGAAATTAATTTTATTGTCCCTGAAATGCTCT
XC              TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAACAGGGAAATTGTTTTTATTGTGCCTGAAATGCTCT
XLS             TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGAAATTATTTTTATTGTCCCTGAAATGCTCT
XM              TGATTTGGAAGCGAAAATCCTTGAAGCAAAGCAAAAGGTTGGTAGCAGGGTAATTGTTTTTATTGTCTCTGA--------
                ═══════════════════════
                          EXON 15

                          * ** ********** ************* *************** ******* **** *** *** *
XLL1            TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL2            TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL3            TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL4            TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL5            TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL6            TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL7            TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL8            TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL9            TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL10           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL11           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL12           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL13           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL14           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL15           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL16           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL17           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL18           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL19           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL20           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL21           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL22           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL23           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XLL24           TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACATTCTATTGGATAACA
XA              TTATCACAGGAAACGTT-AAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACGTTCTATTGGATAGCA
XC              TTATCACATGAAACGTTTAAGGCAATTTGCACTGAAATTATTGCAGTGGTGTCACTGCTTACACGTTCTGTTGGATAACA
XLS             TTATCACAGTAAACGTTTAAGGCAATTTGCACTGAAATTATGGCAGTGGTGTCACTGCTTACACGTTCTATTGGATAACA
XM              -------------CATTTAAGGCAATTTTCACTGAAATTATGGCAGTGGTGTCACTACTTACACATTCTGTTGAATAAC-
```

```
                *** *  * * *** **** ** ***********  *** *  **  ** ** ***
XLL1        ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL2        ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL3        ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL4        ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL5        ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL6        ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL7        ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL8        ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL9        ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL10       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL11       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL12       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL13       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL14       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL15       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL16       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL17       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL18       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL19       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL20       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL21       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL22       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL23       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XLL24       ATGCATTTTCAGCTGTTGGCTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTCCAAAAATACTAAGGTAAAG
XA          ATGCATTTACAGCTGTTGGCTTAATTGCCTAGATACAGCCATACATGCAAAGCTGTAGCTCCAAAGATACGAAGGTAAAT
XC          ATGCATTTTCAGCTGTTGGTTTAATTGCCTAAATACAGCCATATATGCAAAGCTAAAGCTCCAAAGACACTAAGGTAAAG
XLS         ATGCATTTTCAGCTGTTCGTTTAGTTGCCTAGATACAGCCATATATGCAAAGCTATAGCTC-AAAAATACTAAGGTAAAG
XM          ----------CAGATACTGGTTTAATTGCATACATACAGCCATATTTGCCACACTATAGATCCAAAA--------------

            **  **    ** * ******* ***  *** ********* * **** *****  ****    ***   **** *****
XLL1        TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL2        TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL3        TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL4        TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL5        TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL6        TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL7        TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL8        TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL9        TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL10       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL11       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL12       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL13       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL14       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL15       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL16       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL17       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL18       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL19       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL20       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL21       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL22       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL23       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XLL24       TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACAAAGGA
XA          TAAAAACCTTAACAGGTTTCTTCAGATAGAACTATCCTCCAGTATAAAAATAACACTTGCTTTGCCCCATAAACAAAGGA
XC          TAAAAACCTTAAAAGGTTTCTTCAGATAGAAATATCCTCCACTATAAAAATAACATTTGC----CCCATAAAACAAAGGA
XLS         TAAAAACCTTAAAAGGTTTCTTCAGATAGAACTATCCTCCACTATAAAAATAACACTTGC----CCCATAAAACCAAGGA
XM          TAGGAATAGGAAGATGTTTCTTAAGAGGGAAATATCCTCCACTTTAAAGATAACTCTTGC----CCCATAAAACAAAGGA
```

```
                **** ***     * *
XLL1        GATAATCCTCACTTAAA-------------------------------------------------------------
XLL2        GATAATCCTCACTTAAA-------------------------------------------------------------
XLL3        GATAATCCTCACTTAAA-------------------------------------------------------------
XLL4        GATAATCCTCACTTAAA-------------------------------------------------------------
XLL5        GATAATCCTCACTTAAA-------------------------------------------------------------
XLL6        GATAATCCTCACTTAAA-------------------------------------------------------------
XLL7        GATAATCCTCACTTAAA-------------------------------------------------------------
XLL8        GATAATCCTCACTTAAA-------------------------------------------------------------
XLL9        GATAATCCTCACTTAAA-------------------------------------------------------------
XLL10       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL11       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL12       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL13       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL14       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL15       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL16       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL17       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL18       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL19       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL20       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL21       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL22       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL23       GATAATCCTCACTTAAA-------------------------------------------------------------
XLL24       GATAATCCTCACTTAAA-------------------------------------------------------------
XA          GATAATCCTCACTTTAGGCTAGGGCTACATGTGCGTTTTTTGTGCGTCTCGACGCGCTGCGTCAAAAATAATGTTAAAGA
XC          GATAACCCTCTCATAAATTGGATGTATACCTTCCATAGAACTCTACTATTGCTGTCTGCTCGGTTAAATAGGTGGTCCAA
XLS         GATAATCCTCACTTAAAGGAATTGTTCAGTGTAAAAATAAAAACTGGGTCAATAGACAGGCTGTGCAAAAAAAAAATGTT
XM          TATAATCCTAATTTAAATGGGATGTATACCTTCCATAGAATTTGACTGTTGCTCTGCCAGCTCTCCTAACTAGGTGGCCC


XLL1        -------------------------------------------------------------------------------
XLL2        -------------------------------------------------------------------------------
XLL3        -------------------------------------------------------------------------------
XLL4        -------------------------------------------------------------------------------
XLL5        -------------------------------------------------------------------------------
XLL6        -------------------------------------------------------------------------------
XLL7        -------------------------------------------------------------------------------
XLL8        -------------------------------------------------------------------------------
XLL9        -------------------------------------------------------------------------------
XLL10       -------------------------------------------------------------------------------
XLL11       -------------------------------------------------------------------------------
XLL12       -------------------------------------------------------------------------------
XLL13       -------------------------------------------------------------------------------
XLL14       -------------------------------------------------------------------------------
XLL15       -------------------------------------------------------------------------------
XLL16       -------------------------------------------------------------------------------
XLL17       -------------------------------------------------------------------------------
XLL18       -------------------------------------------------------------------------------
XLL19       -------------------------------------------------------------------------------
XLL20       -------------------------------------------------------------------------------
XLL21       -------------------------------------------------------------------------------
XLL22       -------------------------------------------------------------------------------
XLL23       -------------------------------------------------------------------------------
XLL24       -------------------------------------------------------------------------------
XA          TAGAATGTCGCATCGCGTCGCAGCGTCGTGTCCGACGCGACGCGACTGTCGCATGCGGACGCAGC---------------
XC          ACCTAGAGGGCTAGAGGGTAAGAGGGATAGAGGGCACCCCGATGAGAATATTACCTACAAGAAAGAAATGGGCAACTGGT
XLS         TCTAATATAGTTAGTTAGCCAAAAATGTAATGTATAAAGGCTGGAGTGATTTGATGTATATTATGTCAGTCAGATCACTT
XM          AAACCTTGAGAGCTAGAGGGCCATGGGAATCTTACCAACATAAAATAAAATGGGCAGCATGTGGCAGCAGCATTATTAAT
```

```
XLL1    ------------------------------------------------------------------------
XLL2    ------------------------------------------------------------------------
XLL3    ------------------------------------------------------------------------
XLL4    ------------------------------------------------------------------------
XLL5    ------------------------------------------------------------------------
XLL6    ------------------------------------------------------------------------
XLL7    ------------------------------------------------------------------------
XLL8    ------------------------------------------------------------------------
XLL9    ------------------------------------------------------------------------
XLL10   ------------------------------------------------------------------------
XLL11   ------------------------------------------------------------------------
XLL12   ------------------------------------------------------------------------
XLL13   ------------------------------------------------------------------------
XLL14   ------------------------------------------------------------------------
XLL15   ------------------------------------------------------------------------
XLL16   ------------------------------------------------------------------------
XLL17   ------------------------------------------------------------------------
XLL18   ------------------------------------------------------------------------
XLL19   ------------------------------------------------------------------------
XLL20   ------------------------------------------------------------------------
XLL21   ------------------------------------------------------------------------
XLL22   ------------------------------------------------------------------------
XLL23   ------------------------------------------------------------------------
XLL24   ------------------------------------------------------------------------
XA      ------------------------------------------------------------------------
XC      TGGCAGGGCTAAAATGGGCCTAGTTACACTTGTAGAGAATAAGTAAAAGCAGCATACTAATGTTAGTAGAATAATATTAA
XLS     CCTGCTTTTCAGCTCTCTTGGTTTACACTGACTGGTTACCCTGGCTACCAGGCAGTAACCAATCAGAGACTTGAGGGGGA
XM      CCCATTAGTTATGTTCTCTGTTATGAGTACAATTACTTGCAGGACTAAAATGTAGATGATAAGTAAAAGCAGAAATTCTC


XLL1    ------------------------------------------------------------------------
XLL2    ------------------------------------------------------------------------
XLL3    ------------------------------------------------------------------------
XLL4    ------------------------------------------------------------------------
XLL5    ------------------------------------------------------------------------
XLL6    ------------------------------------------------------------------------
XLL7    ------------------------------------------------------------------------
XLL8    ------------------------------------------------------------------------
XLL9    ------------------------------------------------------------------------
XLL10   ------------------------------------------------------------------------
XLL11   ------------------------------------------------------------------------
XLL12   ------------------------------------------------------------------------
XLL13   ------------------------------------------------------------------------
XLL14   ------------------------------------------------------------------------
XLL15   ------------------------------------------------------------------------
XLL16   ------------------------------------------------------------------------
XLL17   ------------------------------------------------------------------------
XLL18   ------------------------------------------------------------------------
XLL19   ------------------------------------------------------------------------
XLL20   ------------------------------------------------------------------------
XLL21   ------------------------------------------------------------------------
XLL22   ------------------------------------------------------------------------
XLL23   ------------------------------------------------------------------------
XLL24   ------------------------------------------------------------------------
XA      ------------------------------------------------------------------------
XC      TTAAATGATGAATACTTTTTTAAAATATTACAAATGAATAGTAAAAGGCTCACAAA------------------------
XLS     GCCACATGGGTCATATCTGTTGCTTTTGAATCTGAGCTGAATGCTGAGGATCAATTGCAAACTCACTGAACAGAAATGTC
XM      ATTTTATTATTATTGTTTTATTATTATTATTATATAAATTAAATCATTAATATATTTAATAATGATTATACAATTCA
```

```
XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    ----------------------------------------------------------------------------------------
XLL5    ----------------------------------------------------------------------------------------
XLL6    ----------------------------------------------------------------------------------------
XLL7    ----------------------------------------------------------------------------------------
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   ----------------------------------------------------------------------------------------
XLL12   ----------------------------------------------------------------------------------------
XLL13   ----------------------------------------------------------------------------------------
XLL14   ----------------------------------------------------------------------------------------
XLL15   ----------------------------------------------------------------------------------------
XLL16   ----------------------------------------------------------------------------------------
XLL17   ----------------------------------------------------------------------------------------
XLL18   ----------------------------------------------------------------------------------------
XLL19   ----------------------------------------------------------------------------------------
XLL20   ----------------------------------------------------------------------------------------
XLL21   ----------------------------------------------------------------------------------------
XLL22   ----------------------------------------------------------------------------------------
XLL23   ----------------------------------------------------------------------------------------
XLL24   ----------------------------------------------------------------------------------------
XA      ----------------------------------------------------------------------------------------
XC      ----------------------------------------------------------------------------------------
XLS     CCATGTGGCCCCCCTTCAAGTCGCTGACTAGCTCAGAGTTATAGAGCTGAAAAGCAGGAAGTTGGATTCTGTCTGTTTTA
XM      AATATGAATATTAATTTTAATAATCATTATACAATTCAAATATAATAGTAAAAGGTTCACAAATGT----------------


XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    ----------------------------------------------------------------------------------------
XLL5    ----------------------------------------------------------------------------------------
XLL6    ----------------------------------------------------------------------------------------
XLL7    ----------------------------------------------------------------------------------------
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   ----------------------------------------------------------------------------------------
XLL12   ----------------------------------------------------------------------------------------
XLL13   ----------------------------------------------------------------------------------------
XLL14   ----------------------------------------------------------------------------------------
XLL15   ----------------------------------------------------------------------------------------
XLL16   ----------------------------------------------------------------------------------------
XLL17   ----------------------------------------------------------------------------------------
XLL18   ----------------------------------------------------------------------------------------
XLL19   ----------------------------------------------------------------------------------------
XLL20   ----------------------------------------------------------------------------------------
XLL21   ----------------------------------------------------------------------------------------
XLL22   ----------------------------------------------------------------------------------------
XLL23   ----------------------------------------------------------------------------------------
XLL24   ----------------------------------------------------------------------------------------
XA      ----------------------------------------------------------------------------------------
XC      ----------------------------------------------------------------------------------------
XLS     TTAGACATCTGTTCACTCCAGCCTTTATGCATTACATTTTTGGCTAACTAACTATATTAGAAATATTTTTTATTTTGCAC
XM      ------------------------------------------------------------AAATGTAATAGTTATTGGCC
```


XLL1 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL2 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL3 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL4 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL5 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL6 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL7 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL8 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL9 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL10 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL11 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL12 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL13 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL14 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL15 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL16 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL17 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL18 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL19 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL20 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL21 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL22 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL23 ----------TGGAATGTATACCTTCCATAGAACTCT
XLL24 ----------TGGAATGTATACCTTCCATAGAACTCT
XA ----------GCGGGATGTATACCTTCCATAGAACTCT
XC ----------TGTATT-AGTGTTTGGCCCCAATAGA---TT
XLS AGCCTATCTATTTACCCAGTTTTTATTTTCACACTGAACTATTCCTTTAAATGGAATGTATACCTTCCATAGAACTTTAC
XM CCAATAGATAACCCATAGCTATATATTTTATTTACTTTAACTTACTTTAACTTTAGAGCACAAGGCTTAAGTTCAGG-CT
XLL1 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL2 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL3 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL4 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL5 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL6 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL7 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL8 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL9 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL10 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL11 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL12 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL13 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL14 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL15 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL16 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL17 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL18 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL19 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL20 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL21 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL22 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL23 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XLL24 ACTATTGCTCTGTTAG--CTCTCTTAAATAGGTGGCCCAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XA ATTATTGCTCTGTTAG--CGCTTTTAAATAGGTGGCCAAAAC--CTAGAGGG--------GAAGAGGGC--------ATC
XC ACTGACA-TCTGTTTA--TTTTATTTACTTTGTCTTTAGAGC--C--GAAGACTT----AGATTCAGCCCTTGGGGGAGC
XLS TATTGCTCTGTTAG----CTCTCTTAAATAGGTGGCTGAAAC--CTAGAGGGCTAGAGGAGAAGAGGGCTAGAGGGCACC
XM ACCCTTGGGGAGATAGAATTTTCTCAGACCACCCCCCTCCTTTGCTCATGTGCTTAATAATATTGACATGCTCTTATCAC

```
                  *       *                *                                  *  **    *  *
XLL1            CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL2            CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL3            CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL4            CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL5            CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL6            CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL7            CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL8            CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL9            CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL10           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL11           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL12           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL13           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL14           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL15           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL16           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL17           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL18           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL19           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL20           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL21           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL22           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL23           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XLL24           CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAGA
XA              CCAGTGAGAATCCTACCTACAAGAATGAAAT-------------------------GGGCCTATTTATACTCGTAGAGA
XC              TAAACAATTCTCAGACCCCTTTTGCTTATTTTAA-----GGTTCCTAA---TATATGAAATATGGTT-TATTAGTCACTC
XLS             CCAATGAGAATCTTACCTACAAGAATGAAATAGGCAACTGATTGGCAGGGCTAAAATGGGCCTAGTTATACTTGTAGAG-
XM              TTATGATGAATACTGTTTTGCATA--TACACAAAGAGGTGCTTTTCACATTTAATCTGCATCTGTTTTGGTTTCTAGATA

                      *         *          **   *  *    *               ** ***         *   **
XLL1            AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL2            AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL3            AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL4            AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL5            AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL6            AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL7            AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL8            AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL9            AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL10           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL11           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL12           AGAAGTAAAAGCAGATATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL13           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL14           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL15           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL16           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL17           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL18           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL19           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL20           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL21           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL22           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL23           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XLL24           AGAAGTAAAAGCAGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XA              ATAAGAAAAAGCAGACATTCTAATGTTAGTAGAATAATATTACTTAAATGATGAATCATTTTAATAATAGTACACATTAA
XC              TCTTATAGTTTCAGGAATAACAAAGGCAGTAAATAAGCATTCACCTCCAAATATCACATTTTGGT--CCACACA-ATTAC
XLS             --AAGTAAAAGCGGACATTCTAATGTTAGTAGAATAATATTAATTAAATGATGAATCATTTTAATAATATTACAAATTAA
XM              CTGTACACTACAAAACACACCT---TCAGATGAATATATATATATATATGGAACAGCACTTTA------TCAAGGATCCT
```

```
                 *       *   * *                     *    *     * *          *  *   *
XLL1        TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL2        TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL3        TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL4        TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL5        TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL6        TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL7        TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL8        TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL9        TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL10       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL11       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL12       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL13       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL14       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL15       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL16       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATACCTGTTTATTT-TATTTACTT---TGCC
XLL17       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL18       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL19       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL20       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL21       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL22       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL23       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XLL24       TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XA          TAACAAAAAGCTCACAAATGTATTAGTACCTGGCCCCACTAGATTACTTATATCTGTATATTT-TATTTACTT---TGCC
XC          AAGTGAGCTAGTGATATA-GTACTGACATGCTCTTTCACTT-ATGATGAATACTGTTTTGTATATACAGAAATGGGTGCT
XLS         TAGCAAAAAGCTCACAAATGTATAAGTACCTGGCCCCACTAGATTACCTATATCTGTTTATTT-TATTTACTT---TGCC
XM          TGAAAAAGGTTCCAATAATGACCAAATGTTGGATACAGC---GTGTCCTAAAATAAATCACTTTAAATGAAGA----AAC

                *  *       *    *   *            *    *    *  * *                          *** *    *
XLL1        TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL2        TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL3        TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL4        TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL5        TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL6        TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL7        TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL8        TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL9        TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL10       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL11       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL12       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL13       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL14       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL15       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL16       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL17       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL18       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL19       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL20       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL21       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL22       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL23       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XLL24       TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGACCCCCTTTGCT--TATCTGGGGT
XA          TTTCGAGCTGATGACTTAGATTCAGGCTGCCATTTGGGGTGCTAAACATTTCTCAGACCCTCTTTGCT--TATCTGGGGT
XC          TTTTACATTTAACAGTCTGAATTTGTTTTGGTTTACAGATACTGTACACTACACTCCTTTAGATGAATA-TATATAAAAT
XLS         TTTAGAGCCGAAGACTTAGATTCAGGCTGCCATTTGGGGAGCTAAACATTTCTCAGGCCCCCTTTGCT--TATCTGGGGT
XM          TGTAATGCTGATCCCTTCTAAATATATATACCTGTGTGTAGATATATATGTTTTATTTATAATCTATACATATATATAAT
```

```
                **      * ** ***   ** *****  * **** **    * ******* *** **          ** * * * * *
XLL1        CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL2        CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL3        CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL4        CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL5        CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL6        CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL7        CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL8        CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL9        CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL10       CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL11       CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL12       CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL13       CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL14       CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL15       CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL16       CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL17       CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL18       CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL19       CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL20       CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL21       CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL22       CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL23       CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XLL24       CTACTTTGCT-TATCTGGG-TCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGGTTGGGTGTTTGATTTCA
XA          CTACTTTGCT-TATCTGGTGTCTACTTAGTAAGTGACCTTAGTAAGTACTTTTAACGCTGTGGTTGGCTGTTTGATTTCA
XC          ATACAGGAGTATACCTGGGTTCAACTTAGTATGTGAACT--CGAAGTACTTTTAACCCTGTGGTTGGGTGTTTGATTTAA
XLS         CTACTTTGCT-TATCTGGGGTCTACTTAATATGTGACCT-AGCA-GTACTTTTAACCCTGTGTTTGGGTGTTTGATTTCA
XM          ATATACAAGTATACCTGTGTTCTACTTAGAATGTGACCT-AACA-GTACTTT-AACCCT---------TGCTAGTTCTTA

                  * **         ***        * *   * *    ***  * *  ***  ***        ** * **    *   *
XLL1        ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL2        ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL3        ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL4        ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL5        ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL6        ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL7        ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL8        ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL9        ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL10       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL11       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL12       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL13       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL14       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL15       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL16       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL17       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL18       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL19       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL20       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL21       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL22       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL23       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XLL24       ATAATATGAGTTACATAT--TAGGAGAATGACATAACTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XA          ATAATATGAGTTATATATATTAGGAGAATGACATAAAGGGCCTCTTCTGTCATAAACATCATTACTTTTGTTAAATACTA
XC          TTAATATCAGATACTTAT--TAGGGGAATGACATAAAGGGCTGCTTCTGTCATAAACAACATGACTTTTGGTAAA---TA
XLS         ATAATATGAGTTATATAT--TAGGAGAATGACATACCTGGCCTCTTCTGTCATAAACAACATTACTTTTGTTAAATACTA
XM          GGGAAATGG------TAT----TAAGGACTGCTT--CTGGCAACATGAGTCCAAAAATATGACACATCTGAAGCAGATTC
```

```
          *       * *  *  ********  * * ***       ****     *        * ***  * ***** ** ***
XLL1      TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL2      TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL3      TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL4      TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL5      TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL6      TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL7      TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL8      TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL9      TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL10     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL11     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL12     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL13     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL14     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL15     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL16     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL17     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL18     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL19     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL20     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL21     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL22     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL23     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XLL24     TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XA        TGTCACTGAAACATTTATGTAAAACGCCT-GGAATTATCACTATGCTTTTTTAGGCTCCCACCGATGGCCAAAGGCTAAT
XC        TGTCACTAAAATATGAATGTAAAATACTTAGGAATTATCACTAT-----TCCAGGCTCCTACCACTGGCCAAAGGGTAAT
XLS       TGGCACTAAAACATTAATGTAAAACACCT-GGAATTATCACTATGCTTTTCCAGGCTCCCACCGCTGGCCAAAGGGTAAT
XM        TTAT----AGACTTGAATGTAAAACACCT-GGA-------CTATGCTTTTT-------CCACCACTAGCCAAGGGTTAAA

          *  ***  ******* ***  *****     ** ****** *  *****    ****  *  ** **      * *****
XLL1      A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL2      A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL3      A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL4      A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL5      A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL6      A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL7      A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL8      A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL9      A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL10     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL11     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL12     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL13     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL14     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL15     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL16     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL17     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL18     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL19     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL20     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL21     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL22     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL23     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XLL24     A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTAGTACCCCATGGAGCTTGAAGCA
XA        A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGCGAGCA----TTTTGTTATCCAATGAAGCTTAAAGCA
XC        A--CAAGTTCTTGTTTGGCGATTTACATCAGTGGGGGCTGCTGTGAGCA----TTTTGGTGCCCCATGGAGTTTAAAGCA
XLS       A--CAAGTTCTTGTTTGGCGATTTACATCTATGGGGGCTGTTGTGAGCA----TTTTGGTACCCCATGGAGCTTGAAGCA
XM        AAACAAAGTCTTGTTAGGCATTTTACTAATATGCGGGCTGCTATGAGCACTTTTTTTGATGCCCCAT------TAAAGCA
```

```
          **    **** * ** *** ****** ** **** * ** ***       ***  ***** *****  * ** ***** *
XLL1      CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL2      CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTT------AATGATGTCTAATGAACCTTTGGCATATAAA
XLL3      CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTT------AATGATGTCTAATGAACCTTTGGCATATAAA
XLL4      CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL5      CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL6      CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL7      CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTT------AATGATGTCTAATGAACCTTTGGCATATAAA
XLL8      CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTT------AATGATGTCTAATGAACCTTTGGCATATAAA
XLL9      CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTT------AATGATGTCTAATGAACCTTTGGCATATAAA
XLL10     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTT------AATGATGTCTAATGAACCTTTGGCATATAAA
XLL11     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL12     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL13     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL14     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL15     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL16     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTT------AATGATGTCTAATGAACCTTTGGCATATAAA
XLL17     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL18     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTT------AATGATGTCTAATGAACCTTTGGCATATAAA
XLL19     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL20     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTT------AATGATGTCTAATGAACCTTTGGCATATAAA
XLL21     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTT------AATGATGTCTAATGAACCTTTGGCATATAAA
XLL22     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL23     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTTGTGTTTAATGATGTCTAATGAACCTTTGGCATATAAA
XLL24     CA--GAGTAGACCATATTCCACTACACTTTTGGTCACTGGTTT------AATGATGTCTAATGAACCTTTGGCATATAAA
XA        CA--AAGTAGACCATATTCCACTACACTTTTGGTTATTGGTTTGTGTTTAATGAAGTCTAATGAACTTTTGGCATATAAA
XC        CACAAAGTAGGCCATATTCCACTACACTTTTGGTTACTGGTTTGTGTTTAATGAAGTCTAATGAACCTT-GGCATATAAA
XLS       CA--AAGTAGACCATATTCTACTACATTTTTGGTCACTGGTTTGTGTTTTATGAAGTCTAATGAACCTTTGGCATATACA
XM        CACAATGTAGGCAATTTTCCACTACAGTTCTGGTTAATGATTTGTGGTTAATGGAGTCTACTGAACCCTTGGTATATAAA

          * ******** **  *                                    *   ****** ** *** ******
XLL1      TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL2      TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL3      TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL4      TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL5      TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL6      TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL7      TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL8      TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL9      TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL10     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL11     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL12     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL13     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL14     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL15     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL16     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL17     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL18     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL19     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL20     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL21     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL22     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL23     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XLL24     TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XA        TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAAGGACCATGCATGCTTCTTCTGCTGCCCAATAAAGAC------
XC        TGTCTTGTTAAGTGCAACTTATACAACTCTATATCAAATGACCATGCATGCTCCTTCTGATGTCCACTAAAGACAAATAT
XLS       TGTCTTGTTAAGTGCAATTTATACAACTCTATATCAAATGACCATGCATGTTTCTTCTGCTGCCCACTAAAGACA--TGT
XM        TATCTTGTTAGGTTTA---------------------------------GCCTCTTCTGCTGGCCACTAAAGATA--TAT
```

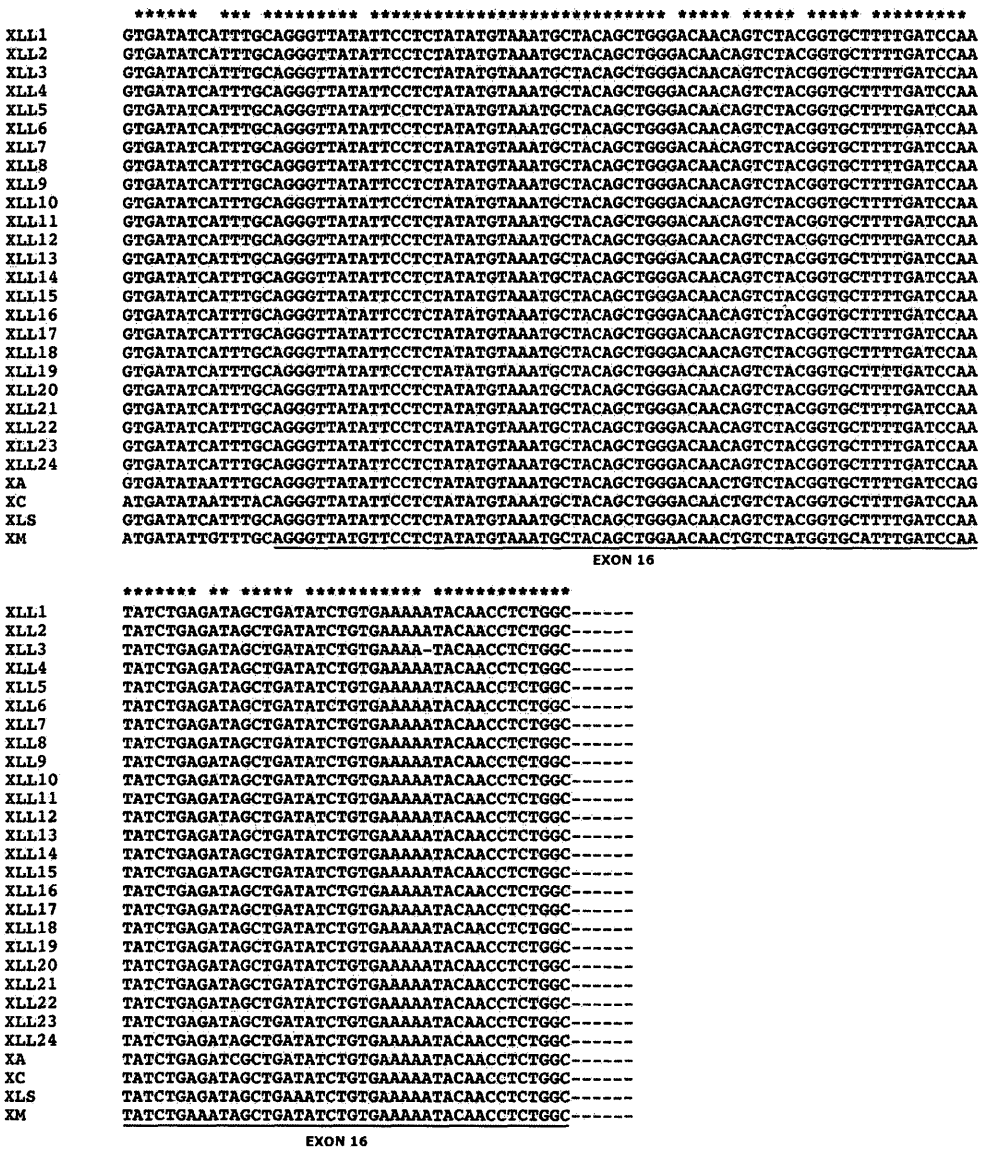
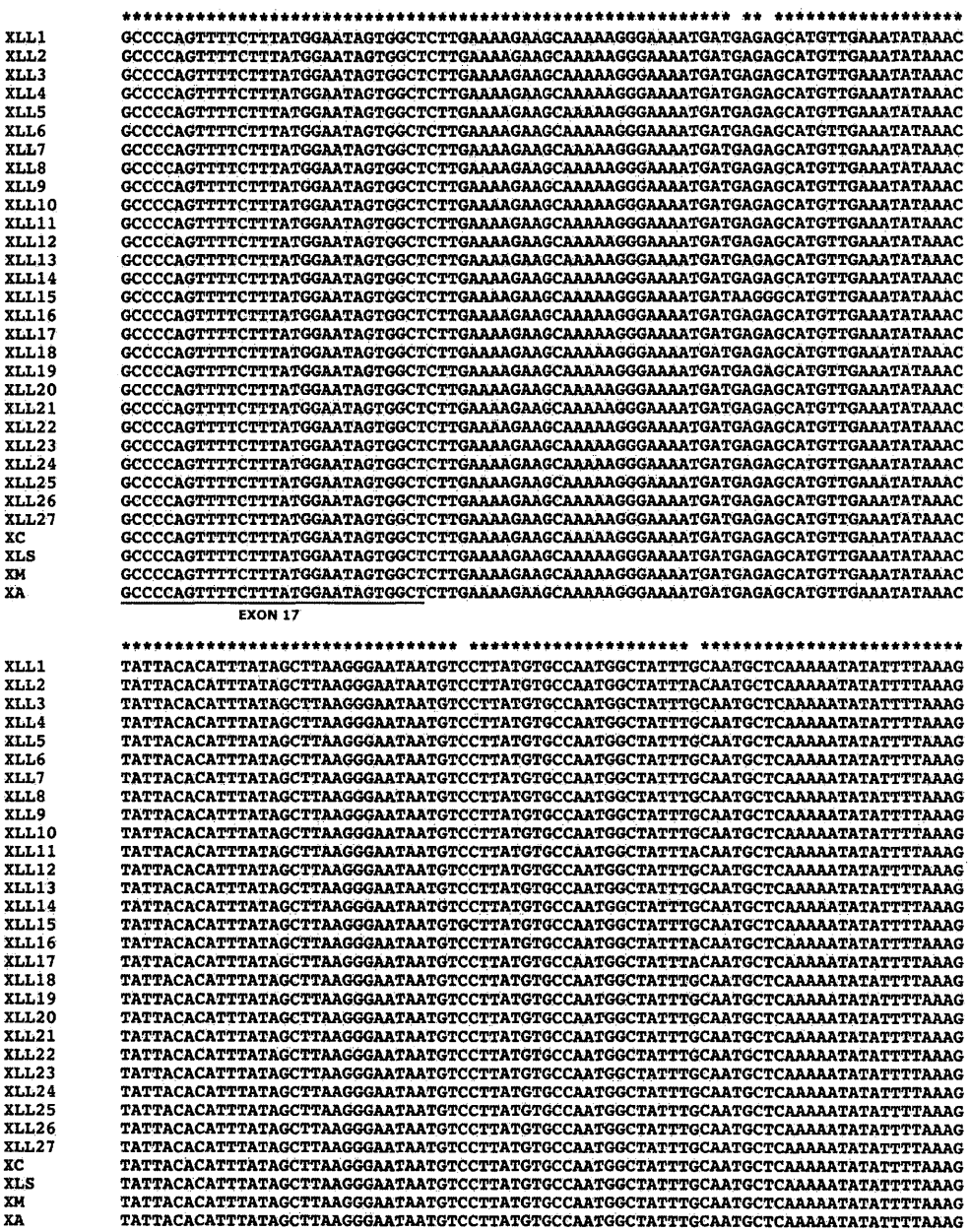


Figure 2. Manually adjusted alignment of *xGAD67* intron 15. A ClustalX 2.0.11 alignment, performed on 24 *X. l. laevis* alleles (XLL1 – XLL24) and four closely related outgroup species (XA, *X. amieti*; XC, *X. clivii*; XLS, *X. l. sudanensis*; XM, *X. muelleri*) under defaults parameters, was modified using Mesquite in order to minimize the number of reported gaps. The regions corresponding to exons 15 and 16 are underlined and labeled. Sequences proceed from 5' to 3' and conserved sites are indicated by an asterisk (*).


XLL1
XLL2
XLL3
XLL4
XLL5
XLL6
XLL7
XLL8
XLL9
XLL10
XLL11
XLL12
XLL13
XLL14
XLL15
XLL16
XLL17
XLL18
XLL19
XLL20
XLL21
XLL22
XLL23
XLL24
XLL25
XLL26
XLL27
XC
XLS
XM
XA
EXON 17

```
                ************************** ******************************************     ******
XLL1            GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL2            GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL3            GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL4            GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL5            GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL6            GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL7            GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL8            GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL9            GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL10           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL11           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL12           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL13           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL14           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL15           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATA-----TCTCTC
XLL16           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL17           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL18           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL19           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL20           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL21           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL22           GGCCACTACAATTGTCCTGCAGAATTATGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL23           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL24           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL25           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL26           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLL27           GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XC              GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XLS             GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XM              GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC
XA              GGCCACTACAATTGTCCTGCAGAATTGTGTCATACCAGAGAATACTGCTGCTGCTATTGTTCTATTATACAGTATCTCTC

                *********** ****************** *********** ************************************
XLL1            TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL2            TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTACTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL3            TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL4            TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL5            TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL6            TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL7            TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL8            TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL9            TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL10           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL11           TGTATGGGCTACGAGAAAACAAAGAGTGCTCTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL12           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL13           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL14           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL15           TGTATGGGCTATGAGAAAACAAAGAGTGCTCTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL16           TGTATGGGCTACGAGAAAACAAAGAGTGCTCTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL17           TGTATGGGCTACGAGAAAACAAAGAGTGCTCTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL18           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL19           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL20           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL21           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL22           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL23           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL24           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL25           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL26           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLL27           TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XC              TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XLS             TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XM              TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
XA              TGTATGGGCTACGAGAAAACAAAGAGTGCTGTGCTGGTTACTGCTACGACAAGAGTATTACATGGCGGGCCCCAGAAATT
```

```
          *** ****** ********************************** ********************************
XLL1      GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL2      GCCTTGAGATATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL3      GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL4      GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL5      GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL6      GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL7      GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL8      GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL9      GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL10     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL11     GCCTTGAGATATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATTTGGTAA
XLL12     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL13     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL14     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL15     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGGTTTTTGGAAGACATGAACAAAAAGGTGATTTGGTAA
XLL16     GCCTTGAGATATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATTTGGTAA
XLL17     GCCTTGAGATATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATTTGGTAA
XLL18     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL19     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL20     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL21     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL22     GCCCTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL23     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL24     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL25     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL26     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLL27     GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XC        GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XLS       GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XM        GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------
XA        GCCTTGAGAGATCTTGTGACCAGACCACCTGATAACAATGGAGATTTTTGGAAGACATGAACAAAAAGGTGATT------

                                  ********************************* ***************** ******
XLL1      --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL2      --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL3      --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL4      --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL5      --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL6      --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL7      --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL8      --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL9      --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL10     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL11     AAAACTAGGGAATTCAAAGAACAATTCTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL12     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL13     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL14     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL15     AAAACTAGGGAATTCAAAGAACAATTCTCGCTGCTCATCTCCTAGAATGAAGGTCATATTGTGCTCATCCAACTCAGCTT
XLL16     AAAACTAGGGAATTCAAAGAACAATTCTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL17     AAAACTAGGGAATTCAAAGAACAATTCTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL18     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL19     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL20     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL21     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL22     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACCCAGCTT
XLL23     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL24     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL25     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL26     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLL27     --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XC        --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XLS       --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XM        --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
XA        --------------------------CTCGCTGCTCATCTCCTAGAATGAAGGTCAAATTGTGCTCATCCAACTCAGCTT
```

```
                ************************************ ********** ******************************
XLL1            GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL2            GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL3            GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL4            GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL5            GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL6            GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL7            GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL8            GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL9            GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL10           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL11           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGG
XLL12           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL13           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL14           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL15           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCCTAATTCTGATTAAGTTATATATAAAAAAAGAAGATATTATGA
XLL16           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGG
XLL17           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGG
XLL18           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL19           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL20           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL21           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL22           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL23           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL24           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL25           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL26           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLL27           GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XC              GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XLS             GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XM              GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA
XA              GTATAACTATCAGACCTTCCACTACCAAGAAATTGGCTTAATTCTGATGAAGTTATATATAAAAAAAGAAGATATTATGA

                ********************************** *********************************************
XLL1            CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL2            CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL3            CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL4            CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL5            CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL6            CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL7            CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL8            CCAGTTTCTGATCAAACCATAGTGACATGCAGACGCTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL9            CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL10           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL11           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL12           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL13           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL14           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL15           CCAGTTTCTGATCAAACCATAGTGACATGCAGACATTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL16           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL17           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL18           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL19           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL20           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL21           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL22           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL23           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL24           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL25           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL26           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLL27           CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XC              CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XLS             CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XM              CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
XA              CCAGTTTCTGATCAAACCATAGTGACATGCAGACGTTTTCGCTTTTCAGCCCTATATATACTTTCAAAGGACAGGGTAAT
```

Figure 3. ClustalX 2.0.11 alignment of *xGAD67* intron 17. Sequences from 27 *X. l. laevis* alleles (XLL1 – XLL27) and four closely related outgroup species (XA, *X. amieti*; XC, *X. clivii*; XLS, *X. l. sudanensis*; XM, *X. muelleri*) were aligned under defaults parameters. The regions corresponding to exons 17 and 18 are underlined and labeled. Sequences proceed from 5' to 3' and conserved sites are indicated by an asterisk (*).

Nucleotide Polymorphism

Alignment of the *xGA67* upstream region in 27 *X. l. laevis* individuals revealed 46 segregating sites and 28 singletons, which are segregating sites that are only represented once in a sample (Table 1). Intron 15 contained only two segregating sites both of which were singletons (Table 2). Nineteen segregating sites were identified in intron 17 with 15 singleton sites (Table 3). Average nucleotide diversity (π) was at least three fold higher in the upstream region (π = 8.84) in comparison to intron 15 and 17, respectively, π = 0.17 and 2.10. Watterson's estimator θ_W for the upstream region was 0.0123 per site while θ_H ranged from 2.78 to 9.62. Intron 15 showed a marked reduction in both θ estimators, except for θ_H along the *X. clivii* lineage. θ_W at the intron 17 locus was equally reduced compared to the upstream region, yet θ_H demonstrated intermediate values between the upstream region and intron 15. Comparison of θ_H across all three loci reveals that the upstream region contains comparatively more variants at higher frequency that two intronic loci, an indicator of genetic hitchhiking.

Indel Polymorphism

Traditional estimators of population diversity only account for the presence and frequency of single nucleotide variants. Indels are common structural changes within the genome and are generated by DNA replication errors, transposon activity, and unequal crossing over during meiosis. Estimators of indels, analogous to the aforementioned nucleotide estimators, have been developed to capture data missed by single nucleotide surveys. The overall pattern of indel polymorphism between the upstream region of

xGAD67 and both introns was similar to that of nucleotide polymorphism; the upstream, presumptive promoter region, is measurably more variable than the corresponding neural proxies (Table 1). The upstream region contains four segregating simple indels with an π_{indel} = 0.65 and Watterson's θ_{indel} = 1.04. Simultaneously accounting for both intron frequency and length, Balhoff and Wray's (BW) measure of average indel length polymorphism was estimated to be π_{BW} = 5.1. Segregating simple indels for intron 15 locus ranged from 1-3. Average pairwise indel diversity, π_{indel}, varied from 0.08 to 1.12 while π_{BW} = 1.53 across all *Xenopus* lineages. Intron 17 had reduced intron diversity across all measures with only one segregating simple indel[2], π_{indel} = 0.0741 and π_{BW} = 0.78 which would propose that the intron itself may be under purifying selection or tightly linked to a locus, possibly within the exon, that has undergone a selective sweep(s).

[2] Measurements of "simple indels" count the number of indels and do not take into account the length of any given indel.

Statistic	Value
Nucleotide Polymorphism	
Population Sample Size	27
Sequence Length	1308
Informative Length	973
k	46
Singletons	28
Watterson's θ_W per site	0.0123
π	8.84 ($p < 0.15$)
Tajima's D_T	-0.98 ($p < 0.165$)
Simple Indel Polymorphism	
K_{indels}	4
Watterson's θ_{indel}	1.04
p indels	0.65
Tajima's D_{indels}	-0.99
Fu and Li's D_{indels}	-1.01
p_{BW}	5.10
p_{BW} per site	0.0039

Other Statistics	*X. l. sudanenesis*	*X. amieti*	*X. clivii*	*X. muelleri*
Substituted Sites (L)	12	1	3	3
Hudson's ρ per site	0.00227	0.00274	0.00264	0.00274
θ_H	9.62 ($p < 0.47$)	5.93 ($p < 0.18$)	2.78 ($p < 0.017$)	5.93 ($p < 0.18$)
Fu and Li's D_{FL}	-2.80 ($p < 0.007$)	-2.80 ($p < 0.007$)	-2.80 ($p < 0.007$)	-2.80 ($p < 0.007$)
Fay and Wu's H	-0.79($p < 0.290$)	2.91($p < 0.53$)	6.06 ($p < 0.134$)	2.91($p < 0.53$)

Table 1. Statistics for the upstream region of *xGAD67*. The amount and pattern of variation was analyzed with various summary statistics. The statistical significance of our results was assessed by coalescent simulation of a 973 bp locus, ρ=0 and k=46, evolving neutrally in panmictic population at stationary. Substituted sites (L) is the number of differences between the respective outgroup and the *X. l. laevis* population sequences minus polymorphic sites.

Statistic	*X. l. sudanenesis*	*X. amieti*	*X. clivii*	*X. muelleri*
Nucleotide Polymorphism				
Population Sample Size	24	24	24	24
Sequence Length	1888	1888	1888	1888
Informative Length	1322	1281	1293	1190
k	2	2	2	2
Singletons	2	2	2	2
Watterson's θ_W per site	0.000405	0.000418	0.000414	0.000450
π	0.17	0.17	0.17	0.17
Tajima's D_T	-1.51	-1.51	-1.51	-1.51
Simple Indel Polymorphism				
K_{indels}	3	3	2	1
Watterson's θ_{indel}	0.80	0.80	0.54	0.27
π_{indels}	1.12	1.12	0.60	0.08
Tajima's D_{indels}	1.00	1.00	0.27	-1.16
Fu and Li's D_{indels}	-0.24	-0.24	-0.72	-1.65
π_{BW}	1.53	1.53	1.53	1.53
π_{BW} per site	0.000808	0.000808	0.000808	0.000808
Other Statistics				
Substituted Sites (L)	50	81	286	379
Hudson's ρ per site	0.00436	0.00436	0.00582	0.00291
θ_H per site	0.00725	0.00725	1.92	0.00362
Fu and Li's D_{FL}	-2.27	-2.27	0.60	-1.65
Fay and Wu's H	0.16	0.16	-1.83	0.08

Table 2. Statistics for the *xGAD67* intron 15. Statistical summary of the amount and pattern of variation in intron 15, a neutral proxy for the *xGAD67* locus. Twenty-four alleles of *X. l. laevis* were aligned with four *Xenopus* outgroups. The inclusion of each outgroup with the population sequences changed the alignment and the number of informative sites. The results from each alignment are therefore provided for each outgroup. The paucity of segregating sites did not permit statistical inferences to be made from modeling the locus by coalescent simulation.

Statistic	
Nucleotide Polymorphism	Value
Population Sample Size	27
Sequence Length	692
Informative Length	654
k	19
Singletons	15
Watterson's θ_W per site	0.00754
π	2.10
Tajima's D_T	-2.03 ($p < 0.006$)

Indel Polymorphism	
k_{indels}	1
Watterson's θ_{indel}	0.26
π_{indels}	0.0741
Tajima's D_{indels}	-1.15
Fu and Li's D_{indels}	-1.69
π_{BW}	0.78
π_{BW} per site	0.00113

Other Statistics	*X. l. sudanenesis*	*X. amieti*	*X. clivii*	*X. muelleri*	
Substituted Sites (L)	0	0	0	0	
Hudson's ρ per site	0.00125	0.00125	0.00125	0.00125	
θ_H per site	0.21	0.21	0.21	0.21	
Fu and Li's D_{FL}	-3.65	-3.65	-3.65	-3.65	($p<0.0005$)
Fay and Wu's H	1.90	1.90	1.90	1.90	($p<0.7$)

Table 3. Statistics for the *xGAD67* intron 17. Statistical summary of the amount and pattern of variation in intron 17, a neutral proxy for the *xGAD67* locus. Twenty-seven alleles of *X. l. laevis* were aligned with four *Xenopus* outgroups. The statistical significance of our results was assessed by coalescent simulation of a 654 bp locus, ρ=0 and k=19, evolving neutrally in panmictic population at stationary.

Estimating Hudson's ρ and test statistic distributions of a stable panmictic population

Recombination, or more accurately the rate of recombination, is frequently used in population and evolutionary genetics to interpret polymorphism data. The scaled recombination rate $\rho = N\eta\mu$ was estimated using Hudson's maxhap program, as described in the Methods section and Appendix, for each outgroup species. Estimates of ρ across all four outgroup species at all three loci suggest that recombination is not pervasive (Tables 1-3).

Coalescence theory has become indispensable for interpreting polymorphism data and the likelihood of demographic or selective forces in shaping patterns of variation (Rosenberg and Nordborg, 2002). Coalescent simulations were conducted with Hudson's ms program to generate 100,000 replicate samples of a stable, neutrally evolving panmictic population with no recombination. The distribution of π, Tajima's *D*, Fu and Li's *D*, Fay and Wu's *H* test statistics were extracted with Hudson's sample_stats program in order to determine the likelihood of our results given a neutral model (Fig. 4). The sampling distributions of the all test statistics were not appreciably altered even at the highest estimated rate of recombination, $\rho = 0.00582$, therefore, null distributions with and without estimated rates of recombination are equally conservative.

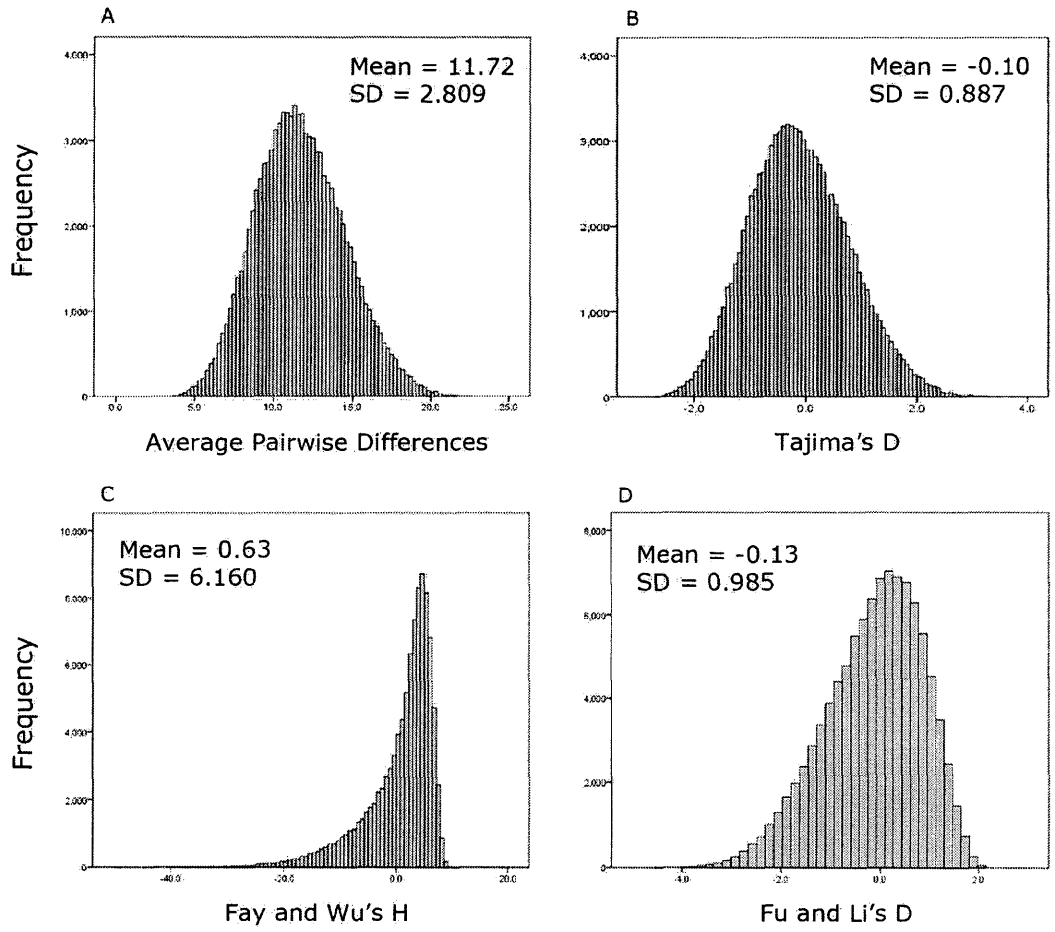


Figure 4. Distribution of test statistics for a neutrally evolving population modeled by coalescent simulation. A genealogical history with a specified, Poisson distributed number of mutations was generated for a locus equal to the size of the upstream region of *xGAD67*. Hudson's ms generated 100,000 replicate sample populations and the values of four summary statistics (A-D) were extracted for each replicate sample with Hudson's sample_stats program. The observed statistic for the upstream region was calculated for each distribution to be (A) 8.84, (B) -0.98, (C) 2.77, and (D) 2.80. Summary statistics for C-D are averaged across all four *Xenopus* outgroups.

Selection Analysis

The significance of sequence variation at the *xGAD67* locus was analyzed to determine if the immediate upstream region has been subject to selection. Four test statistics were employed on a *X. l. laevis* population sample of 27 alleles and four closely related species in order to detect the presence of selection. Tajima's D_T and Fay and Wu's H utilize different θ estimators, while Fu and Li's D_{FL} assesses the genealogical distribution of polymorphisms against a neutral expectation. In contrast, the HKA test relies on a neutral proxy. We chose to use intron 15 as the neutral proxy for the HKA test based on the observed number of within species differences and between species substitutions. Interestingly, no substituted sites were identified in the outgroup for intron 17. These statistics, except for the HKA test, were developed to detect selection from the frequency spectrum of polymorphic sites in a population sample and therefore indels are excluded from the analysis. Subsequently, D_T and D_{FL} tests were modified and reported as D_{INDEL} and D_{INDEL} in order to accommodate the possibility that indels could be under selection. In order to investigate the likely variation contained in summary statistics of nucleotide variation, we determined the polymorphism frequency spectrum for D_T and D_{FL} was also determined along the axis of transcription in a sliding window frame of 101 bases with a 10 base slide. Essentially, each summary statistic is calculated for a small region (e.g. 101 bps.). The 101 bp window or region of interest is then shifted 10 bps along the sequence and the summary statistics are again calculated

Fu and Li's *D* test statistic suggests that the promoter region of *xGAD67* contains an excess of rare alleles that is significantly greater than the level expected in a neutrally evolving population (Table 1). The distribution of extreme low frequency alleles, summarized as negative Fu an Li values, is particularly concentrated in the most proximal region immediately upstream of the *xGAD67* coding region, approximately 175 bps upstream from the transcriptional start site (Fig 5). In this region, the Fu and Li statistic takes on the most negative values for the upstream region, indicating an abundance of rare alleles (i.e. singletons) immediately upstream of the transcriptional start site. Using Fu and Li's nomenclature, there is an abundance of new mutations in the external branches of the *Xenopus* genealogy (i.e. *X. l. laevis* clade) compared to the older internal branches (i.e. *Xenopus* outgroups). An excess of low frequency alleles can be generated by rapid population expansion/contraction, positive selection, or purifying selection. Positive Fay and Wu *H* values found in three out of four outgroup species for all three loci suggest that positive selection is unlikely as the signature of genetic hitchhiking[3] would preferentially increase θ_H to produce negative *H* values (Table 1). Our inference is strengthened by the very similar, and positive, distribution of *H* values immediately upstream of the transcriptional start for all four *Xenopus* outgroups (Figure 6). Moreover, dramatic perturbations in population size would also generate negative Tajima's *D* values, which were shown to be statistically consistent with a stable panmictic population in the upstream region. Therefore, a model of purifying selection in the region immediately upstream of *xGAD67* is consistent with our results.

[3] Genetic hitchhiking occurs with a locus has recently undergone a strong selective sweep and risen to high frequency within the population because of positive, favorable selection for a new variant. Nearby, neutral variation, also rises or hitchhikes to high frequency because of linkage.

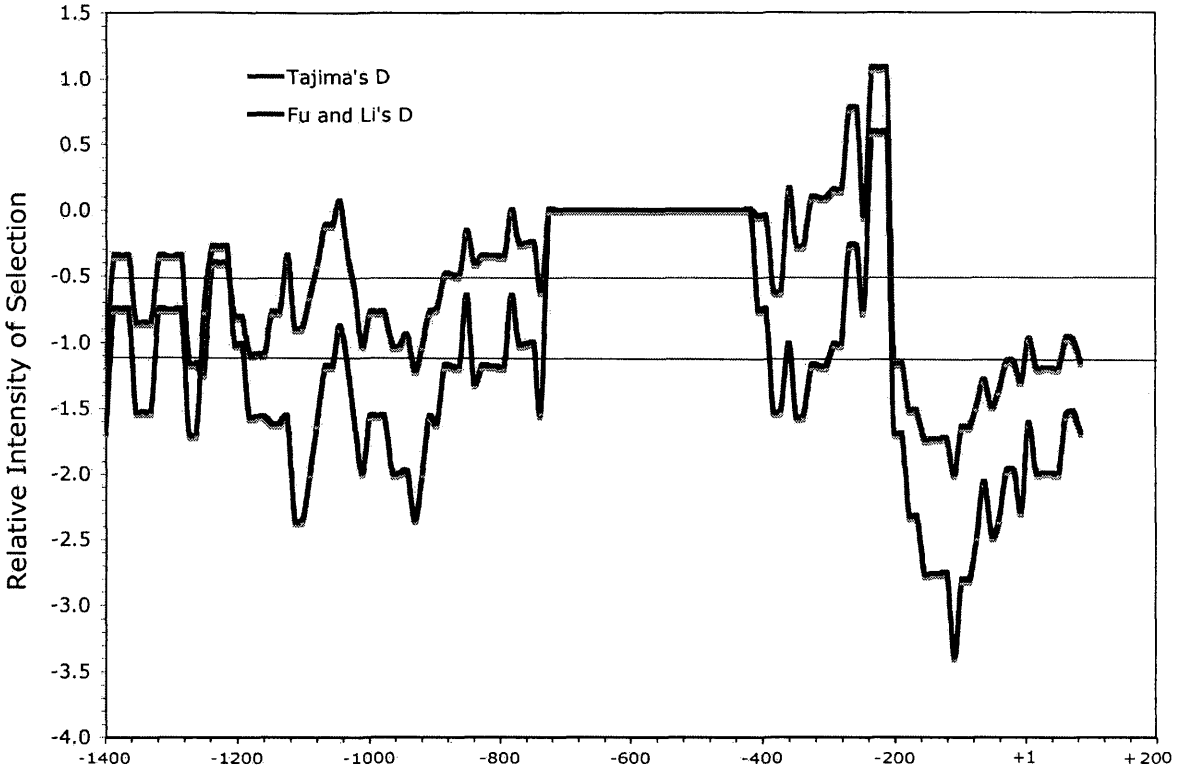


Figure 5. Polymorphism frequency spectrum for the upstream region of *xGAD67*. Tajima's *D* (D_T) and Fu and Li's *D* (D_{FL}) test statistics are plotted along the axis of transcription from 5' to 3' in 101 bp sliding window frame with a slide of 10 bps. The red and blue horizontal lines indicate the average D_T and D_{FL} values, respectively. The +1 indicates the transcriptional start site.

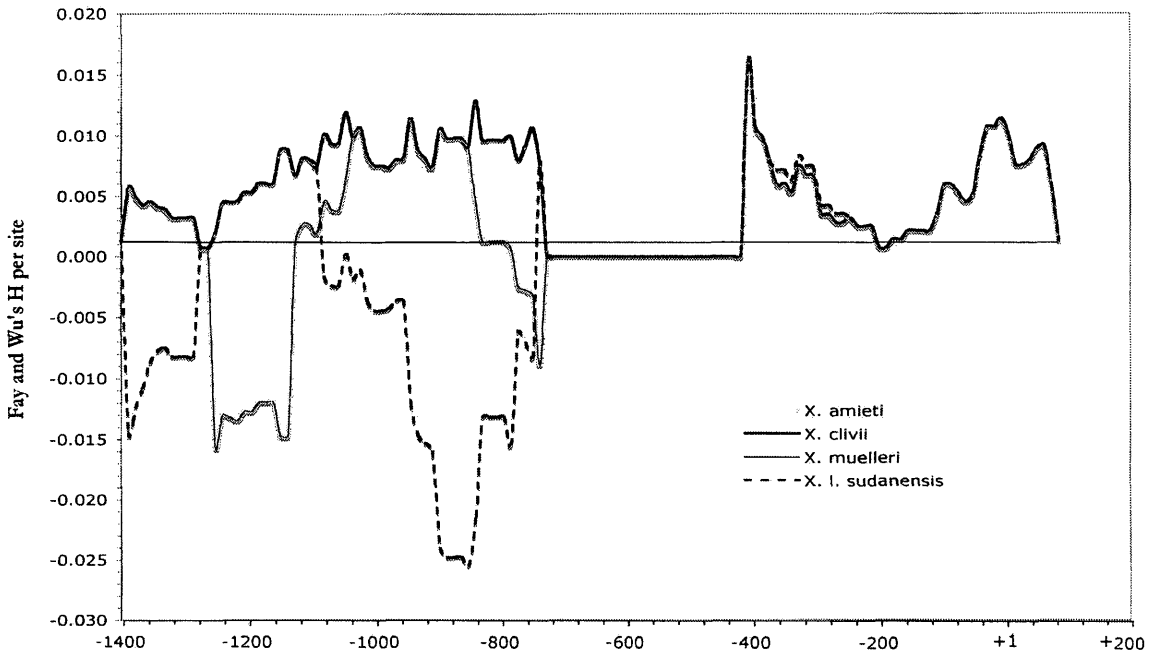


Figure 6. Fay and Wu's *H* site-frequency spectrum of the upstream region of *xGAD67*. Fay and Wu's *H* statistic was calculated per site in 101 bp sliding window frame (10 bp slide) and plotted against the axis of transcription from 5' to 3'. The black horizontal line indicates the average *H* value per site for all four *Xenopus* species. The +1 indicates the transcriptional start site.

Incongruously, intron 17 also displays a strong signature of purifying selection (Table 3). Fu and Li's *D* values are substantially higher than the expectation of a similarly sized locus evolving neutrally in a stable panmictic population ($p < 0.0005$). Tajima's *D* are in concordance with directional, potentially purifying, selection ($p< 0.006$), but do not discount the possibility of population expansions as a confounding scenario. While not statistically significant, Fay and Wu's *H* of 1.9 suggest that positive selection is unlikely.

Statistical inference of population parameters for intron 15 alone is problematic due to the deficiency of segregating sites within the *X. l. laevis* population. Modeling the distribution of any given test statistic by coalescence assumes that number of mutations Poisson distributed along a genealogy can be approximated by the number of observed mutations at a locus. However, when the number of polymorphic sites is very low then mutations cannot be randomly distributed and any statistical inference based on our test statistics are inconclusive (Depaulis et al., 2001). On a qualitative level, there is a positive correlation between the number of substitutions (L) and the evolutionary distance of *Xenopus* outgroups (Table 2). The linear accumulation of substitutions in each of the outgroups combined with a dearth of polymorphic sites may suggest balancing selection (Fu and Li, 1993), however, the paucity of informative sites makes such a conclusion speculative.

Results from the HKA test also support a substantial deviation from a neutral model of evolution (Table 5). Intron 15 was chosen as a neutral proxy to investigate the upstream region of *xGAD67* because it contained both within species polymorphism and between species divergence. The ratio of within species to between species variation was considerably greater for the upstream region (3.83) compared with intron 15 (0.04). Moreover, the degree of variation within the *X. l. laevis* population was 31.25x higher for the promoter than intron

15 while the level of divergence was only 0.32 (Table 4). These relationships of within species variation to between species divergence for the upstream region and intron 15 are consistent across all the outgroup species.

Species		Upstream region	Intron 15	Ratio (upstream/intron 15)
X. l. sudanenesis				
	Within species	0.047	0.002	31.250
	Between species	0.012	0.038	0.326
	Ratio (within/between)	3.833	0.040	
X. amieti				
	Within species	0.047	0.002	31.250
	Between species	0.001	0.061	0.017
	Ratio (within/between)	46.000	0.025	
X. clivii				
	Within species	0.047	0.002	31.250
	Between species	0.003	0.216	0.014
	Ratio (within/between)	15.333	0.007	
X. muelleri				
	Within species	0.047	0.002	31.250
	Between species	0.003	0.287	0.011
	Ratio (within/between)	15.333	0.005	

Table 4. Comparison of nucleotide variation in the upstream region and intron 15 of *xGAD67*. The amount of polymorphism and divergence in the 973 bp upstream region and 1,322 bp intron 15 is reported within a population of 24 *X. l. laevis* alleles and between four closely related *Xenopus* species.

Region	Polymorphic Sites	Substituted Sites			
		X. l. sudanenesis	*X. amieti*	*X. clivii*	*X. muelleri*
Promoter	46	12	1	3	3
Intron 15	2	50	81	286	379
		(p < 3.49E-06)	(p <1.52E-11)	(p <1.88E-38)	(p <0)

Table 5. HKA test of neutrality for the upstream region of *xGAD67*. Intron 15 was chosen as a neutral proxy for the nearby upstream region of *xGAD67*. The HKA test of a neutral model assumes that the level of polymorphism and divergence are products of the neutral mutation rate and therefore equivalent. The statistical significance of the amount of variation in 24 *X l. laevis* alleles compared with four closely related *Xenopus* species is reported.

DISCUSSION AND FUTURE DIRECTIONS

We have provided evidence that the proximal promoter region, immediately upstream of *xGAD67*, may be subject to purifying selection. More distal elements, approximately 800 bps upstream, may be under lineage specific selection. Changes within noncoding regulatory elements can have significant impacts on phenotypic diversity. Noncoding elements regulating the expression of developmental genes have been shown to drive major morphological transitions. However, whether the logic of regulatory evolution for large scale morphological features extends to the cellular physiological level remains unclear. Moreover, the intractability of identifying *cis*-regulatory elements has slowed progress in understanding regulatory evolution in comparison to protein coding regions. Significant progress has been made in elucidating the relationships between transcript abundance and patterns of noncoding variation in many model systems (Johnson et al., 2009;

Kim et al., 2009; Lawniczak et al., 2008; Tung et al., 2009; Zhang and Borevitz, 2009). Detailed studies in sea urchins have been able to superimpose selection analysis with maps of experimentally identified *cis*-regulatory elements at loci of several terminal differentiation genes (Balhoff and Wray, 2005; Walters et al., 2008). However, such analyses are difficult with most vertebrates and not feasible in higher primates. The relevance of *cis*-regulatory elements identified in cell assays to biological processes in the organism remains questionable.

Xenopus has continued to be a major model system for early vertebrate embryology with advances in genome sequencing and transgene technology (Ogino and Ochi, 2009). The ability to assay for *cis*-regulatory driven gene expression *in vivo* for a vertebrate system is significant. Progress in understanding the speciation genetics of African clawed frogs (Evans, 2008) coupled with experimental tractability make *Xenopus* a powerful model for investigating regulatory evolution.

We have provided an evolutionary genetic analysis of the upstream region of *xGAD67* using a South African population of *X. l. .laevis* individuals and four closely related *Xenopus* species: *X. amieti, X. clivii, X. l. sudanensis,* and *X. muelleri.* We have provided evidence that ~1.3 kbps upstream of the translational start site may have been subject to purifying selection, and therefore, contain functional regulatory elements. The pattern of extremely low frequency alleles suggest that the intensity of purifying selection is particularly concentrated in the -200 to +150 region were putative *cis*-regulatory elements may be recognizing retina specific transcription factors necessary for GABAergic differetnation. This region is capable of driving retina-specific *xGAD67* expression when cloned into a green fluorescent protein (GFP) reporter and monitored in a developing transgenic *X. l.*

laevis embryo (data unpublished). The pattern of endogenous retina-specific *xGAD67* expression is widely conserved in vertebrates and the action of purifying selection is not unexpected. However, similar studies in sea urchins have shown that widespread sequence conservation is not a necessary prerequisite driving conserved patterns of expression (Oda-Ishii et al., 2005). The features of *cis*-regulatory regions, such as modularity, are theoretically capable of emerging from nonadaptive processes (Lynch, 2007a; Lynch, 2007b) and redundancy in regulatory elements has been shown for promoters with shadow enhancers[4] (Wray and Babbitt, 2008). However, the suggestion that gene expression is maintained by TF binding site turnover, where sites are lost and gained concomitantly, has found little empirical support (Doniger and Fay, 2007).

Our study suggests that if regulatory elements are present in the region 1.3 kbps upstream of *xGAD67* then developmental expression of this locus is intolerant of mid to high frequency variation. Alternatively, the excess of rarest alleles may be due to sequencing or PCR amplification errors, however we attempted to mitigate those sources of error by using multiple clones and bi-directional sequencing to determine the sequence of each allele. Moreover, the abundance of low frequency alleles was not observed at other intronic loci. Another possibility is that our primers may have been allele specific, or allele degenerate whereby allopolyploid paralogs may have been included in the analysis. Both of these situations seem unlikely based on our initial electrophoretic results of PCR products where only 1 or 2 bands were observed (data not shown). Also, there exist the possibility where *xGAD67* locus may be subject to an unusual pattern of inheritance for allopolyploids where multivalents form during meiosis (Sammut et al., 2002).

---

[4] *Cis*-regulatory elements that recognize a similar complement of transcription factors acting on the main regulatory elements.

Based on the polymorphism frequency spectrum, we propose that the region -200 to +150 of *xGAD67* likely contains *cis*-regulatory elements. In order to determine if this region or others nearby contain regulatory elements, a variety of transgene constructs will need to be generated. The significance of the polymorphism frequency spectrum of *xGAD67* requires an *in vivo* assay in order functionally characterize the regulatory elements. Transgenic *X. laevis* embryos can be generated by sperm nuclear injection of upstream regions fused to a GFP reporter (Smith et al., 2006). A variety of transgenic methods have been developed for high transgene efficiency (e.g. REMI, *ISce*I meganuclease, or transposon methods) (Ogino and Ochi, 2009). Constructs should be generated for both the longest and shortest allele in the population, 1,252 and 982 bps in length, respectively. Insertional elements on the order of 200 bps may have arisen from transposon activity and could contain entire regulatory elements with enhancer or silencer activity. The -200 to +150 bp region will need to be assayed with a basal promoter-driven GFP reporter plasmid that contains *cis* elements necessary for recruiting the basal transcriptional machinery needed for activity in all cells.

Experimental identification of *cis*-regulatory elements can be further resolved using a variety of techniques to investigate specific TF binding targets. For example, *xDistal-less-4* (*xDll-4*) and *xGAD67* expression has been shown to overlap in the forebrain (Brox et al., 2003). *xDll-4* binding sites could be identified *in silico* and independently, or coordinately, disrupted through site directed mutagenesis. High resolution regulatory maps could be superimposed on the site-frequency spectrum to further resolve which sites are evolving under selection. Lastly, the endogenous pattern of *xGAD67* expression should also be assessed in each of the outgroups. While distant vertebrates are known to express *xGAD67*

similarly during development, *in situ* hybridization should be performed to confirm that lineage specific changes in *xGAD67* expression have not occurred.

We find the amount of polymorphic and substituted sites within introns 15 and 17 an unexpected result for a region presumed to be neutral. Intron 15 is an intermediately sized intron at the *xGAD67* locus with an expected pattern of substitutions (Table 2) sustained along lineages of increasing phylogenetic distance (Evans et al., 2004). Curiously, only two segregating sites, both singletons, were identified among 24 alleles. A possibility exists that our original cloning attempts were allele specific whereby a considerable amount of the variation was excluded from our analysis. Amplifying intron 15 with an alternative set of primers would likely resolve whether our initial set was indeed allele specific. However, inspection of the alignment at two indel sites, one and six bps in length, respectively, indicate the presence of at least two different alleles of intermediate frequency isolated by PCR from our population sample.

Conversely, intron 17 contains 19 segregating sites but appears to accumulate no differences along any of our *Xenopus* lineages. Our study relies on one sequence from each outgroup, similar to Balhoff and Wray (2005) and Walters *et al* (2008), and may not be representative. The inverse relationship in variation between introns 15 and 17 is perplexing. Attempts to generalize the relationship between intron length and divergence have produced contradictory results, possibly attributed to lineage specific genome evolution (Gazave et al., 2007; Haddrill et al., 2005).

In summary, the upstream region of *xGAD67* may contain *cis*-regulatory elements that are under selective constraint compared with both the neutral expectation and a nearby neutral region. Since no consensus sequence(s) exists for identifying promoter elements, we

propose that the polymorphism frequency spectrum can provide a detailed map of prospective regulatory elements for genes with conserved expression patterns and functions. The sea urchin literature has several detailed functional descriptions of evolutionary processes acting within promoter regions (Balhoff and Wray, 2005; Walters et al., 2008), while few cases exist in vertebrates. We hope that our analysis will complement the functional utility of *Xenopus* as a vertebrate model system for understanding gene regulation as a major component of the evolutionary process.

APPENDIX

Source code for summary statistics and sliding window analysis

The summary statistics and sliding window analyses were carried our with software written and provided by R.H.. The source code for all of our analyses are provided below. Separate module scripts have been prepared that make performing individual analysis easier; an example is provided later in the Appendix. The following software is known to run with Python 2.5 and Biopython 1.42. The HKA test was implemented using chi-squared distributions, which are implemented in C with a Python wrapper. That code is not provided below.

```
Import Bio.Align.Generic  # Standard.
Import Bio.Alphabet      # "
import Bio.Alphabet.IUPAC # "
import copy              # "
import math              # "
import operator          # "
import re               # "
import sys               # "
import chiSquared        # Custom.

# MyAlignment is a facade to hide the ugliness of Biopython's Alignment.
Class MyAlignment(Bio.Align.Generic.Alignment):
  def __init__(self):
    Bio.Align.Generic.Alignment.__init__(self,
Bio.Alphabet.Gapped(Bio.Alphabet.IUPAC.IUPACAmbiguousDNA(), '-'))
  def addSequence(self, sequence):
    Bio.Align.Generic.Alignment.add_sequence(self, '', sequence)
  def getNucleotides(self, site):
    return list(Bio.Align.Generic.Alignment.get_column(self, site))
  def getSequence(self, taxon):
    return Bio.Align.Generic.Alignment.get_seq_by_num(self, taxon).tostring()
  def getSiteCount(self):
    return Bio.Align.Generic.Alignment.get_alignment_length(self)
  def getTaxonCount(self):
    return len(Bio.Align.Generic.Alignment.get_all_seqs(self))
  def concatenate(self, other):
    result = MyAlignment()
```

```
      for taxon in range(self.getTaxonCount()):
result.addSequence(self.getSequence(taxon)+other.getSequence(taxon))
      return result
   @classmethod
   def concatenation(dummy, first, second):
      result = MyAlignment()
      for taxon in range(first.getTaxonCount()):
result.addSequence(first.getSequence(taxon)+second.getSequence(taxon))
      return result

# FASTA_FileToMyAlignment is clumsy.  A record parser and an iterator would be graceful.
Def FASTA_FileToMyAlignment(   ilename):
   handle = open(   ilename, 'r')
   entries = handle.read().split(">")
   handle.close()
   alignment = MyAlignment()
   for entry in entries[1:]: alignment.addSequence(entry[entry.index("\n")+1:].replace("\n", ""))
   return alignment

# PHYLIP_FileToMyAlignment is clumsy.  A record parser and an iterator would be graceful.
Def PHYLIP_FileToMyAlignment(   ilename):
   handle = open(   ilename, 'r')
   lines = handle.readlines()
   handle.close()
   alignment = MyAlignment()
   for line in lines[1:]:
      matchData = re.search("^\S+\s+(\S+)$", line)
      if matchData != None:
         alignment.addSequence(matchData.group(1))
      else:
         alignment.addSequence(line[10:-2])
   return alignment

# A SND is a single nucleotide difference, which might be a polymorphism within a population or a
substitution between populations or both.
# Snd objects usually aren't directly useful to users; they're meant to be created within Sample objects.
Class Snd:
   def __init__(self, site, nucleotides, populations, sndClasses):
      self.site = site
      self.sndClasses = sndClasses
      self.nucleotides = {}
      self.nucleotideFrequencies = {}
      for population in populations.keys():
         self.nucleotides[population] = [nucleotides[taxon] for taxon in populations[population]]
         self.nucleotideFrequencies[population] = {}
         for nucleotide in set(self.nucleotides[population]):
self.nucleotideFrequencies[population][nucleotide] = self.nucleotides[population].count(nucleotide)
   def __str__(self):
      return "(%s, %s, %s, %s)" % (str(self.site), str(self.nucleotides), str(self.nucleotideFrequencies),
str(self.sndClasses))

# A simple indel has two alleles, one nothing but gaps and the other free of gaps.  Two or more such
features immediately adjacent to each other are a complex indel.
```

```
# SimpleIndel objects usually aren't directly useful to users; they're meant to be created within Sample objects.
Class SimpleIndel:
  def __init__(self, site, size, alleles, populations, indelClasses):
    self.site = site
    self.size = size
    self.indelClasses = indelClasses
    self.alleles = {}
    self.alleleFrequencies = {}
    for population in populations.keys():
      self.alleles[population] = [alleles[taxon] for taxon in populations[population]]
      self.alleleFrequencies[population] = {}
      for allele in set(self.alleles[population]): self.alleleFrequencies[population][allele] = self.alleles[population].count(allele)
  def __str__(self):
    return "(%s, %s, %s, %s, %s)" % (str(self.site), str(self.size), str(self.alleles), str(self.alleleFrequencies), str(self.indelClasses))

# A complex indel is any indel that isn't simple.
# ComplexIndel objects usually aren't directly useful to users; they're meant to be created within Sample objects.
Class ComplexIndel:
  def __init__(self, site, size, alleles, populations, indelClasses):
    self.site = site
    self.size = size
    self.indelClasses = indelClasses
    self.alleles = {}
    self.alleleFrequencies = {}
    for population in populations.keys():
      self.alleles[population] = [alleles[taxon] for taxon in populations[population]]
      self.alleleFrequencies[population] = {}
      for allele in set(self.alleles[population]): self.alleleFrequencies[population][allele] = self.alleles[population].count(allele)
  def __str__(self):
    return "(%s, %s, %s, %s, %s)" % (str(self.site), str(self.size), str(self.alleles), str(self.alleleFrequencies), str(self.indelClasses))

# Sites containing ambiguous nucleotides are ignored apart from optional diplotype expansion.
ambiguousNucleotideSet = set((
  'R', # A, G
  'Y', # C, T
  'W', # A, T
  'S', # C, G
  'M', # A, C
  'K', # G, T
  'B', # C, G, T
  'D', # A, G, T
  'H', # A, C, T
  'V', # A, C, G
  'N'  # A, C, G, T
))

# Synonymy in coding sequences is assessed using the standard genetic code.
aminoAcids = {
```

```
'TTT': 'F', # Phe
'TTC': 'F', # "
'TTA': 'L', # Leu
'TTG': 'L', # "
'TCT': 'S', # Ser
'TCC': 'S', # "
'TCA': 'S', # "
'TCG': 'S', # "
'TAT': 'Y', # Tyr
'TAC': 'Y', # "
'TAA': '.', # Ter
'TAG': '.', # "
'TGT': 'C', # Cys
'TGC': 'C', # "
'TGA': '.', # Ter
'TGG': 'W', # Trp
'CTT': 'L', # Leu
'CTC': 'L', # "
'CTA': 'L', # "
'CTG': 'L', # "
'CCT': 'P', # Pro
'CCC': 'P', # "
'CCA': 'P', # "
'CCG': 'P', # "
'CAT': 'H', # His
'CAC': 'H', # "
'CAA': 'Q', # Gln
'CAG': 'Q', # "
'CGT': 'R', # Arg
'CGC': 'R', # "
'CGA': 'R', # "
'CGG': 'R', # "
'ATT': 'I', # Ile
'ATC': 'I', # "
'ATA': 'I', # "
'ATG': 'M', # Met
'ACT': 'T', # Thr
'ACC': 'T', # "
'ACA': 'T', # "
'ACG': 'T', # "
'AAT': 'N', # Asn
'AAC': 'N', # "
'AAA': 'K', # Lys
'AAG': 'K', # "
'AGT': 'S', # Ser
'AGC': 'S', # "
'AGA': 'R', # Arg
'AGG': 'R', # "
'GTT': 'V', # Val
'GTC': 'V', # "
'GTA': 'V', # "
'GTG': 'V', # "
'GCT': 'A', # Ala
'GCC': 'A', # "
```

```
    'GCA': 'A', # "
    'GCG': 'A', # "
    'GAT': 'D', # Asp
    'GAC': 'D', # "
    'GAA': 'E', # Glu
    'GAG': 'E', # "
    'GGT': 'G', # Gly
    'GGC': 'G', # "
    'GGA': 'G', # "
    'GGG': 'G'  # "
}

def Translate(sequence):
    translation = ""
    site = 0
    while site < len(sequence)-3:
        if "-" in sequence[site:site+3] or sequence[site] in ambiguousNucleotideSet or sequence[site+1] in
ambiguousNucleotideSet or sequence[site+2] in ambiguousNucleotideSet:
            translation += "-"
        else:
            translation += aminoAcids[sequence[site:site+3]]
        site += 3
    return translation

# Sample objects usually aren't directly useful to users; they're meant to be created within Statistics
objects.
Class Sample:
    def __init__(self, alignment, populations = {}, siteClasses = {}, options = {}):
        self.populations = copy.deepcopy(populations)
        self.siteClasses = copy.deepcopy(siteClasses)
        self.options = copy.deepcopy(options)
        if populations == {}:
            self.populations['_all'] = range(alignment.getTaxonCount())
        else:
            self.populations['_all'] = list(set(reduce(operator.concat, populations.values())))
        if siteClasses == {}:
            self.siteClasses['_all'] = range(alignment.getSiteCount())
        else:
            self.siteClasses['_all'] = list(set(reduce(operator.concat, siteClasses.values())))
        if options.has_key('expand_diplotypes'): alignment = self._expandDiplotypes(alignment)
        self.taxonCounts = {}
        for population in self.populations.keys(): self.taxonCounts[population] =
len(self.populations[population])
        self.siteCounts = {}
        for siteClass in self.siteClasses.keys(): self.siteCounts[siteClass] = len(self.siteClasses[siteClass])
        self.unambiguousSiteCounts = {}
        for siteClass in self.siteClasses.keys(): self.unambiguousSiteCounts[siteClass] = 0
        self.informativeSiteCounts = {}
        for siteClass in self.siteClasses.keys(): self.informativeSiteCounts[siteClass] = 0
        self.snds = []
        self.sndClasses = self.siteClasses.keys()
        self.simpleIndels = []
        self.complexIndels = []
        self.indelClasses = self.siteClasses.keys()
```

```
      indelFlag = False
      for site in self.siteClasses['_all']:
        nucleotideList = alignment.getNucleotides(site)
        nucleotideSet = set([nucleotideList[taxon] for taxon in self.populations['_all']])
        if len(nucleotideSet & ambiguousNucleotideSet) > 0: continue
        siteClassesThisSite = []
        for siteClass in self.siteClasses.keys():
          if site in self.siteClasses[siteClass]: siteClassesThisSite.append(siteClass)
        for siteClass in siteClassesThisSite: self.unambiguousSiteCounts[siteClass] += 1
        if '-' in nucleotideSet:
          if not indelFlag:
            indelFlag = True
            indelAlleles = [[nucleotide] for nucleotide in nucleotideList]
            indelClasses = siteClassesThisSite
          else:
            if indelClasses == siteClassesThisSite:
              for taxon in range(alignment.getTaxonCount()):
indelAlleles[taxon].append(nucleotideList[taxon])
            else:
              self._completeIndel(site, indelAlleles, indelClasses)
              indelAlleles = [[nucleotide] for nucleotide in nucleotideList]
              indelClasses = siteClassesThisSite
        else:
          if indelFlag:
            self._completeIndel(site, indelAlleles, indelClasses)
            indelFlag = False
          for siteClass in siteClassesThisSite: self.informativeSiteCounts[siteClass] += 1
          if len(nucleotideSet) > 1: self.snds.append(Snd(site, nucleotideList, self.populations,
siteClassesThisSite))
      else:
        if indelFlag: self._completeIndel(self.siteCounts['_all'], indelAlleles, indelClasses)
      if 'coding' in self.siteClasses.keys():
        self.sndClasses += ['nonsynonymous', 'synonymous']
        if options.has_key('frame'):
          sequenceFrame = options['frame']
        else:
          sequenceFrame = 0
        for snd in self.snds:
          siteFrame = (sequenceFrame+snd.site)%3
          if siteFrame == 0 and snd.site <= self.siteCounts['_all']-3:
            nucleotideList1 = alignment.getNucleotides(snd.site+1)
            nucleotideSet1 = set([nucleotideList1[taxon] for taxon in self.populations['_all']])
            if len(nucleotideSet1 & ambiguousNucleotideSet) > 0 or '-' in nucleotideSet1: continue
            nucleotideList2 = alignment.getNucleotides(snd.site+2)
            nucleotideSet2 = set([nucleotideList2[taxon] for taxon in self.populations['_all']])
            if len(nucleotideSet2 & ambiguousNucleotideSet) > 0 or '-' in nucleotideSet2: continue
            nucleotideList0 = alignment.getNucleotides(snd.site)
            nucleotideSet0 = set([nucleotideList0[taxon] for taxon in self.populations['_all']])
            codonSet = set([nucleotideList0[taxon]+nucleotideList1[taxon]+nucleotideList2[taxon] for
taxon in self.populations['_all']])
            for context in set([codon[1]+codon[2] for codon in codonSet]):
              if len(set([aminoAcids[nucleotide+context[0]+context[1]] for nucleotide in
nucleotideSet0])) > 1:
                snd.sndClasses.append('nonsynonymous')
```

```
                    break
                else:
                    snd.sndClasses.append('synonymous')
            elif siteFrame == 1 and snd.site >= 1 and snd.site <= self.siteCounts['_all']-2:
                nucleotideList0 = alignment.getNucleotides(snd.site-1)
                nucleotideSet0 = set([nucleotideList0[taxon] for taxon in self.populations['_all']])
                if len(nucleotideSet0 & ambiguousNucleotideSet) > 0 or '-' in nucleotideSet0: continue
                nucleotideList2 = alignment.getNucleotides(snd.site+1)
                nucleotideSet2 = set([nucleotideList2[taxon] for taxon in self.populations['_all']])
                if len(nucleotideSet2 & ambiguousNucleotideSet) > 0 or '-' in nucleotideSet2: continue
                nucleotideList1 = alignment.getNucleotides(snd.site)
                nucleotideSet1 = set([nucleotideList1[taxon] for taxon in self.populations['_all']])
                codonSet = set([nucleotideList0[taxon]+nucleotideList1[taxon]+nucleotideList2[taxon] for
taxon in self.populations['_all']])
                for context in set([codon[0]+codon[2] for codon in codonSet]):
                    if len(set([aminoAcids[context[0]+nucleotide+context[1]] for nucleotide in
nucleotideSet1])) > 1:
                        snd.sndClasses.append('nonsynonymous')
                        break
                else:
                    snd.sndClasses.append('synonymous')
            elif siteFrame == 2 and snd.site >= 2:
                nucleotideList0 = alignment.getNucleotides(snd.site-2)
                nucleotideSet0 = set([nucleotideList0[taxon] for taxon in self.populations['_all']])
                if len(nucleotideSet0 & ambiguousNucleotideSet) > 0 or '-' in nucleotideSet0: continue
                nucleotideList1 = alignment.getNucleotides(snd.site-1)
                nucleotideSet1 = set([nucleotideList1[taxon] for taxon in self.populations['_all']])
                if len(nucleotideSet1 & ambiguousNucleotideSet) > 0 or '-' in nucleotideSet1: continue
                nucleotideList2 = alignment.getNucleotides(snd.site)
                nucleotideSet2 = set([nucleotideList2[taxon] for taxon in self.populations['_all']])
                codonSet = set([nucleotideList0[taxon]+nucleotideList1[taxon]+nucleotideList2[taxon] for
taxon in self.populations['_all']])
                for context in set([codon[0]+codon[1] for codon in codonSet]):
                    if len(set([aminoAcids[context[0]+context[1]+nucleotide] for nucleotide in
nucleotideSet2])) > 1:
                        snd.sndClasses.append('nonsynonymous')
                        break
                else:
                    snd.sndClasses.append('synonymous')
        self.gappedLength = self.siteCounts['_all']          # Deprecated.
        Self.ungappedLength = self.informativeSiteCounts['_all'] # "
    def _expandDiplotypes(self, alignment):
        newAlignment = MyAlignment()
        oldToNew = []
        newTaxon = 0
        for taxon in range(alignment.getTaxonCount()):
            sequence = alignment.getSequence(taxon)
            if taxon in self.options['expand_diplotypes']:
                newSequence1 = list(sequence)
                newSequence2 = list(sequence)
                for site in range(len(sequence)):
                    if sequence[site] == 'R':
                        newSequence1[site] = 'A'
                        newSequence2[site] = 'G'
```

```
            elif sequence[site] == 'Y':
              newSequence1[site] = 'C'
              newSequence2[site] = 'T'
            elif sequence[site] == 'W':
              newSequence1[site] = 'A'
              newSequence2[site] = 'T'
            elif sequence[site] == 'S':
              newSequence1[site] = 'C'
              newSequence2[site] = 'G'
            elif sequence[site] == 'M':
              newSequence1[site] = 'A'
              newSequence2[site] = 'C'
            elif sequence[site] == 'K':
              newSequence1[site] = 'G'
              newSequence2[site] = 'T'
          newSequence1 = ''.join(newSequence1)
          newSequence2 = ''.join(newSequence2)
          newAlignment.addSequence(newSequence1)
          newAlignment.addSequence(newSequence2)
          oldToNew.append([newTaxon, newTaxon+1])
          newTaxon += 2
        else:
          newAlignment.addSequence(sequence)
          oldToNew.append([newTaxon])
          newTaxon += 1
      newPopulations = {}
      for population in self.populations.keys(): newPopulations[population] = reduce(operator.concat,
[oldToNew[taxon] for taxon in self.populations[population]])
      self.populations = newPopulations
      return newAlignment
    def _completeIndel(self, site, indelAlleles, indelClasses):
      indelAlleleList = [''.join(indelAllele) for indelAllele in indelAlleles]
      indelAlleleSet = set([indelAlleleList[taxon] for taxon in self.populations['_all']])
      indelSize = len(indelAlleleList[0])
      if len(indelAlleleSet) == 2 and '-'*indelSize in indelAlleleSet:
        self.simpleIndels.append(SimpleIndel(site-indelSize, indelSize, indelAlleleList, self.populations,
indelClasses))
      else:
        self.complexIndels.append(ComplexIndel(site-indelSize, indelSize, indelAlleleList,
self.populations, indelClasses))

class Statistics:
  def __init__(self, alignment, populations = {}, siteClasses = {}, options = {}):
    self.sample = Sample(alignment, populations, siteClasses, options)
    self.cache = {}
  def getTaxonCount(self, population = '_all'):
    return self.sample.taxonCounts[population]
  def getSiteCount(self, siteClass = '_all'):
    return self.sample.siteCounts[siteClass]
  def getUnambiguousSiteCount(self, siteClass = '_all'):
    return self.sample.unambiguousSiteCounts[siteClass]
  def getInformativeSiteCount(self, siteClass = '_all'):
    return self.sample.informativeSiteCounts[siteClass]
  def getSndCount(self, sndClass = '_all'):
```

```
        sndCount = 0
        for snd in self.sample.snds:
            if sndClass in snd.sndClasses: sndCount += 1
        return sndCount
    def getSimpleIndelCount(self, indelClass = '_all'):
        simpleIndelCount = 0
        for indel in self.sample.simpleIndels:
            if indelClass in indel.indelClasses: simpleIndelCount += 1
        return simpleIndelCount
    def getComplexIndelCount(self, indelClass = '_all'):
        complexIndelCount = 0
        for indel in self.sample.complexIndels:
            if indelClass in indel.indelClasses: complexIndelCount += 1
        return complexIndelCount
    def getIndelCount(self, indelClass = '_all'):
        indelCount = 0
        for indel in self.sample.simpleIndels+self.sample.complexIndels:
            if indelClass in indel.indelClasses: indelCount += 1
        return indelCount
    # K is the number of polymorphic sites (also known as segregating sites) within a population (also
known as S).
    def getK(self, population = '_all', sndClass = '_all'):
        if self.cache.has_key(('K', population, sndClass)):
            return self.cache[('K', population, sndClass)]
        else:
            K = 0
            for snd in self.sample.snds:
                if len(snd.nucleotideFrequencies[population]) > 1 and sndClass in snd.sndClasses: K += 1
            self.cache[('K', population, sndClass)] = K
            return K

    # thetaW is Watterson's estimator of theta (4 N_e u).
    def get_thetaW(self, population = '_all', sndClass = '_all'):
        if self.cache.has_key(('thetaW', population, sndClass)):
            return self.cache[('thetaW', population, sndClass)]
        else:
            a = 0.0
            for i in range(1, self.sample.taxonCounts[population]): a += 1.0/i
            if a > 0.0:
                thetaW = self.getK(population, sndClass)/a
            else:
                thetaW = 0.0
            self.cache[('thetaW', population, sndClass)] = thetaW
            return thetaW
    def get_thetaW_PerSite(self, population = '_all', sndClass = '_all', siteClass = '_all'):
        length = self.sample.informativeSiteCounts[siteClass]
        if length > 0:
            thetaW_PerSite = self.get_thetaW(population, sndClass)/length
        else:
            thetaW_PerSite = 0.0
        return thetaW_PerSite
    # pi is the average number of single nucleotide differences between two sequences from a population
(also known as theta_pi).
    Def get_pi(self, population = '_all', sndClass = '_all'):
```

```
        if self.cache.has_key(('pi', population, sndClass)):
            return self.cache[('pi', population, sndClass)]
        else:
            pi = 0
            n = self.sample.taxonCounts[population]
            if n > 1:
                for snd in self.sample.snds:
                    if sndClass in snd.sndClasses:
                        freqs = snd.nucleotideFrequencies[population]
                        for nuc1 in freqs.keys():
                            for nuc2 in freqs.keys():
                                if nuc1 != nuc2: pi += freqs[nuc1]*freqs[nuc2]
                pi /= 1.0*n*(n-1)
            self.cache[('pi', population, sndClass)] = pi
            return pi
    def get_piPerSite(self, population = '_all', sndClass = '_all', siteClass = '_all'):
        length = self.sample.informativeSiteCounts[siteClass]
        if length > 0:
            piPerSite = self.get_pi(population, sndClass)/length
        else:
            piPerSite = 0.0
        return piPerSite
    # D_T is Tajima's D.
    def getD_T(self, population = '_all', sndClass = '_all'):
        n = self.sample.taxonCounts[population]
        K = self.getK(population, sndClass)
        if n > 1 and K > 0:
            a1 = 0.0
            a2 = 0.0
            for i in range(1, n):
                a1 += 1.0/i
                a2 += 1.0/(i*i)
            b1 = (n+1)/(3.0*(n-1))
            b2 = 2*(n*n + n + 3)/(9.0*n*(n-1))
            c1 = b1 - 1/a1
            c2 = b2 - (n+2)/(a1*n) + a2/(a1*a1)
            e1 = c1/a1
            e2 = c2/(a1*a1 + a2)
            D_T = (self.get_pi(population, sndClass)-self.get_thetaW(population, sndClass))/math.sqrt(e1*K +
e2*K*(K-1))
        else:
            D_T = 0.0
        return D_T
    # K_SimpleIndel is the number of simple indels polymorphic within a population.
    Def getK_SimpleIndel(self, population = '_all', indelClass = '_all'):
        if self.cache.has_key(('K_SimpleIndel', population, indelClass)):
            return self.cache[('K_SimpleIndel', population, indelClass)]
        else:
            K = 0
            for indel in self.sample.simpleIndels:
                if len(indel.alleleFrequencies[population]) > 1 and indelClass in indel.indelClasses: K += 1
            self.cache[('K_SimpleIndel', population, indelClass)] = K
            return K
    def getSingletonCountSimpleIndel(self, population = '_all', indelClass = '_all'):
```

```
    singletonCount = 0
    for indel in self.sample.simpleIndels:
      if len(indel.alleleFrequencies[population]) == 2 and 1 in
indel.alleleFrequencies[population].values() and indelClass in indel.indelClasses: singletonCount += 1
    return singletonCount
  # thetaW_SimpleIndel is Watterson's estimator of theta (4 N_e u) for simple indels.
  Def get_thetaW_SimpleIndel(self, population = '_all', indelClass = '_all'):
    if self.cache.has_key(('thetaW_SimpleIndel', population, indelClass)):
      return self.cache[('thetaW_SimpleIndel', population, indelClass)]
    else:
      a = 0.0
      for i in range(1, self.sample.taxonCounts[population]): a += 1.0/i
      if a > 0.0:
        thetaW = self.getK_SimpleIndel(population, indelClass)/a
      else:
        thetaW = 0.0
      self.cache[('thetaW_SimpleIndel', population, indelClass)] = thetaW
      return thetaW
  def get_thetaW_SimpleIndelPerSite(self, population = '_all', indelClass = '_all', siteClass = '_all'):
    length = self.sample.unambiguousSiteCounts[siteClass]
    if length > 0:
      thetaW_PerSite = self.get_thetaWSimpleIndel(population, indelClass)/length
    else:
      thetaW_PerSite = 0.0
    return thetaW_PerSite
  # piSimpleIndel is the average number of simple-indel differences between two sequences from a
population.
  Def get_piSimpleIndel(self, population = '_all', indelClass = '_all'):
    if self.cache.has_key(('piSimpleIndel', population, indelClass)):
      return self.cache[('piSimpleIndel', population, indelClass)]
    else:
      pi = 0
      n = self.sample.taxonCounts[population]
      if n > 1:
        for indel in self.sample.simpleIndels:
          if indelClass in indel.indelClasses:
            freqs = indel.alleleFrequencies[population]
            for al1 in freqs.keys():
              for al2 in freqs.keys():
                if al1 != al2: pi += freqs[al1]*freqs[al2]
        pi /= 1.0*n*(n-1)
      self.cache[('piSimpleIndel', population, indelClass)] = pi
      return pi
  def get_piSimpleIndelPerSite(self, population = '_all', indelClass = '_all', siteClass = '_all'):
    length = self.sample.unambiguousSiteCounts[siteClass]
    if length > 0:
      piPerSite = self.get_piSimpleIndel(population, indelClass)/length
    else:
      piPerSite = 0.0
    return piPerSite
  # D_T_SimpleIndel is Tajima's D for simple indels.
  Def getD_T_SimpleIndel(self, population = '_all', indelClass = '_all'):
    n = self.sample.taxonCounts[population]
    K = self.getK_SimpleIndel(population, indelClass)
```

```
    if n > 1 and K > 0:
      a1 = 0.0
      a2 = 0.0
      for i in range(1, n):
        a1 += 1.0/i
        a2 += 1.0/(i*i)
      b1 = (n+1)/(3.0*(n-1))
      b2 = 2*(n*n + n + 3)/(9.0*n*(n-1))
      c1 = b1 – 1/a1
      c2 = b2 – (n+2)/(a1*n) + a2/(a1*a1)
      e1 = c1/a1
      e2 = c2/(a1*a1 + a2)
      D_T = (self.get_piSimpleIndel(population, indelClass)-self.get_thetaW_SimpleIndel(population,
indelClass))/math.sqrt(e1*K + e2*K*(K-1))
    else:
      D_T = 0.0
    return D_T
  # piBW is Balhoff and Wray's measure of indel polymorphism.
  Def get_piBW(self, population = '_all', indelClass = '_all'):
    if self.cache.has_key(('piBW', population, indelClass)):
      return self.cache[('piBW', population, indelClass)]
    else:
      piBW = 0.0
      n = self.sample.taxonCounts[population]
      if n > 1:
        for indel in self.sample.simpleIndels+self.sample.complexIndels:
          if indelClass in indel.indelClasses:
            freqs = indel.alleleFrequencies[population]
            for al1 in freqs.keys():
              for al2 in freqs.keys():
                if al1 != al2: piBW += freqs[al1]*freqs[al2]*self._piBW_weight(al1, al2, indel.size)
        piBW /= 1.0*n*(n-1)
      self.cache[('piBW', population, indelClass)] = piBW
      return piBW
  def _piBW_weight(self, al1, al2, size):
    efSizes = []
    efSize1 = 0
    efSize2 = 0
    for site in range(size):
      if al1[site] == '-':
        if al2[site] != '-': efSize1 += 1
      else:
        if efSize1 > 0:
          efSizes.append(efSize1)
          efSize1 = 0
      if al2[site] == '-':
        if al1[site] != '-': efSize2 += 1
      else:
        if efSize2 > 0:
          efSizes.append(efSize2)
          efSize2 = 0
    if efSize1 > 0: efSizes.append(efSize1)
    if efSize2 > 0: efSizes.append(efSize2)
    weight = 0.0
```

```
    for efSize in efSizes: weight += 1.0 + math.log(efSize, 10)
    return weight
  def get_piBW_PerSite(self, population = '_all', indelClass = '_all', siteClass = '_all'):
    length = self.sample.unambiguousSiteCounts[siteClass]
    if length > 0:
      piBW_PerSite = self.get_piBW(population, indelClass)/length
    else:
      piBW_PerSite = 0.0
    return piBW_PerSite
  # L is the number of substituted sites (also known as fixed differences) between two populations (also known as D).
  # Following the McDonald–Kreitman convention, if a site is polymorphic within either population, it isn't substituted.
  Def getL(self, population1, population2, sndClass = '_all'):
    (population1, population2) = self._orderPopulations(population1, population2)
    if self.cache.has_key(('L', population1, population2, sndClass)):
      return self.cache[('L', population1, population2, sndClass)]
    else:
      L = 0
      for snd in self.sample.snds:
        if sndClass in snd.sndClasses:
          freqs1 = snd.nucleotideFrequencies[population1]
          freqs2 = snd.nucleotideFrequencies[population2]
          if len(freqs1) == 1 and len(freqs2) == 1 and freqs1.keys()[0] != freqs2.keys()[0]: L += 1
      self.cache[('L', population1, population2, sndClass)] = L
      return L

  # D_FL is Fu and Li's D.
  # The first population is the ingroup, and the second is the outgroup.
  # Following the DnaSP convention, polymorphic sites in the ingroup that can't be polarized using the outgroup are ignored.
  Def getD_FL(self, population1, population2, sndClass = '_all'):
    n = self.sample.taxonCounts[population1]
    eta = 0
    eta_e = 0
    for snd in self.sample.snds:
      if sndClass in snd.sndClasses:
        freqs1 = snd.nucleotideFrequencies[population1]
        freqs2 = snd.nucleotideFrequencies[population2]
        if len(freqs1) > 1 and len(freqs2) == 1 and freqs2.keys()[0] in freqs1.keys():
          eta += 1
          if freqs1[freqs2.keys()[0]] == n-1: eta_e += 1
    if n > 1 and eta > 0:
      a = 0.0
      b = 0.0
      for i in range(1, n):
        a += 1.0/i
        b += 1.0/(i*i)
      if n == 2:
        c = 1.0
      else:
        c = 2*(n*a – 2*(n-1))/((n-1)*(n-2))
      v = 1 + a*a*(c – (n+1)/(1.0*(n-1)))/(b + a*a)
      u = a-1-v
```

```
        D_FL = (eta – a*eta_e)/math.sqrt(u*eta + v*eta*eta)
      else:
        D_FL = 0.0
      return D_FL
    # F_FL is Fu and Li's F.
    # The first population is the ingroup, and the second is the outgroup.
    # Following the DnaSP convention, polymorphic sites in the ingroup that can't be polarized using the
outgroup are ignored.
    Def getF_FL(self, population1, population2, sndClass = '_all'):
      n = self.sample.taxonCounts[population1]
      pi = 0
      eta = 0
      eta_e = 0
      for snd in self.sample.snds:
        if sndClass in snd.sndClasses:
          freqs1 = snd.nucleotideFrequencies[population1]
          freqs2 = snd.nucleotideFrequencies[population2]
          if len(freqs1) > 1 and len(freqs2) == 1 and freqs2.keys()[0] in freqs1.keys():
            for nuc1 in freqs1.keys():
              for nuc2 in freqs1.keys():
                if nuc1 != nuc2: pi += freqs1[nuc1]*freqs1[nuc2]
            eta += 1
            if freqs1[freqs2.keys()[0]] == n-1: eta_e += 1
      if n > 1 and eta > 0:
        pi /= 1.0*n*(n-1)
        a = 0.0
        b = 0.0
        for i in range(1, n):
          a += 1.0/i
          b += 1.0/(i*i)
        if n == 2:
          c = 1.0
        else:
          c = 2*(n*a – 2*(n-1))/((n-1)*(n-2))
        v = (c + 2.0*(n*n + n + 3)/(9*n*(n-1)) – 2.0/(n-1))/(b + a*a)
        u = (1 + (n+1)/(3.0*(n-1)) – 4.0*(n+1)*(a + 1.0/n – 2.0*n/(n+1))/((n-1)**2))/a – v
        F_FL = (pi – eta_e)/math.sqrt(u*eta + v*eta*eta)
      else:
        F_FL = 0.0
      return F_FL
    # thetaH is Fay and Wu's estimator of theta (4 N_e u).
    # The first population is the ingroup, and the second is the outgroup.
    # Following the DnaSP convention, polymorphic sites in the ingroup that can't be polarized using the
outgroup are ignored.
    Def get_thetaH(self, population1, population2, sndClass = '_all'):
      if self.cache.has_key(('thetaH', population1, population2, sndClass)):
        return self.cache[('thetaH', population1, population2, sndClass)]
      else:
        thetaH = 0.0
        n = self.sample.taxonCounts[population1]
        if n > 1:
          S = {}
          for i in range(1, n): S[i] = 0
          for snd in self.sample.snds:
```

```
            if sndClass in snd.sndClasses:
              freqs1 = snd.nucleotideFrequencies[population1]
              freqs2 = snd.nucleotideFrequencies[population2]
              if len(freqs1) > 1 and len(freqs2) == 1 and freqs2.keys()[0] in freqs1.keys(): S[n-
freqs1[freqs2.keys()[0]]] += 1
          for i in range(1, n): thetaH += S[i]*i*i
          thetaH *= 2.0/(n*(n-1))
        self.cache[('thetaH', population1, population2, sndClass)] = thetaH
        return thetaH
  def get_thetaH_PerSite(self, population1, population2, sndClass = '_all', siteClass = '_all'):
    length = self.sample.informativeSiteCounts[siteClass]
    if length > 0:
      thetaH_PerSite = self.get_thetaH(population1, population2, sndClass)/length
    else:
      thetaH_PerSite = 0.0
    return thetaH_PerSite
  # H is Fay and Wu's H.
  # The first population is the ingroup, and the second is the outgroup.
  # Following the DnaSP convention, polymorphic sites in the ingroup that can't be polarized using the
outgroup are ignored.
  Def getH(self, population1, population2, sndClass = '_all'):
    n = self.sample.taxonCounts[population1]
    theta_pi = 0
    if n > 1:
      S = {}
      for i in range(1, n): S[i] = 0
      for snd in self.sample.snds:
        if sndClass in snd.sndClasses:
          freqs1 = snd.nucleotideFrequencies[population1]
          freqs2 = snd.nucleotideFrequencies[population2]
          if len(freqs1) > 1 and len(freqs2) == 1 and freqs2.keys()[0] in freqs1.keys(): S[n-
freqs1[freqs2.keys()[0]]] += 1
      for i in range(1, n): theta_pi += S[i]*i*(n-i)
      theta_pi *= 2.0/(n*(n-1))
    H = theta_pi-self.get_thetaH(population1, population2, sndClass)
    return H
  def getH_PerSite(self, population1, population2, sndClass = '_all', siteClass = '_all'):
    length = self.sample.informativeSiteCounts[siteClass]
    if length > 0:
      H_PerSite = self.getH(population1, population2, sndClass)/length
    else:
      H_PerSite = 0.0
    return H_PerSite

  # HKApValue is Hudson–Kreitman–Aguade p-value for two "loci".
  # The first SND class is the first "locus", and the second SND class is the second "locus".
  Def getHKApValue(self, population1, population2, sndClass1, sndClass2):
    n1 = self.sample.taxonCounts[population1]
    n2 = self.sample.taxonCounts[population2]
    if n1 > n2:
      population2, population1 = population1, population2
      n2, n1 = n1, n2
    if n1 > 1:
      K11 = 1.0*self.getK(population1, sndClass1)
```

```
    K12 = 1.0*self.getK(population1, sndClass2)
    K1 = K11+K12
    if K1 > 0.0:
      K21 = 1.0*self.getK(population2, sndClass1)
      K22 = 1.0*self.getK(population2, sndClass2)
      K2 = K21+K22
      D1 = 1.0*self.getL(population1, population2, sndClass1)
      D2 = 1.0*self.getL(population1, population2, sndClass2)
      D = D1+D2
      a1 = a2 = b1 = b2 = 0.0
      for i in range(1, n1):
        a1 += 1.0/i
        b1 += 1.0/(i*i)
      for i in range(1, n2):
        a2 += 1.0/i
        b2 += 1.0/(i*i)
      f = (K2/a2)/(K1/a1)
      g = (1.0+f)/2.0
      T = D/(K1/a1) - g
      theta1 = (K11+K21+D1)/(a1 + f*a2 + g + T)
      theta2 = (K12+K22+D2)/(a1 + f*a2 + g + T)
      EK11 = a1*theta1
      EK12 = a1*theta2
      EK21 = a2*theta1
      EK22 = a2*theta2
      ED1 = (g+T)*theta1
      ED2 = (g+T)*theta2
      VK11 = EK11 + b1*theta1**2
      VK12 = EK12 + b1*theta2**2
      VK21 = EK21 + b2*theta1**2
      VK22 = EK22 + b2*theta2**2
      VD1 = ED1 + (g*theta1)**2
      VD2 = ED2 + (g*theta2)**2
      X2 = 0.0
      if VK11 != 0.0: X2 += (K11-EK11)**2/VK11
      if VK12 != 0.0: X2 += (K12-EK12)**2/VK12
      if VK21 != 0.0: X2 += (K21-EK21)**2/VK21
      if VK22 != 0.0: X2 += (K22-EK22)**2/VK22
      if VD1 != 0.0: X2 += (D1-ED1)**2/VD1
      if VD2 != 0.0: X2 += (D2-ED2)**2/VD2
      p = chiSquared.pValue(X2, 2)
    else:
      p = float('nan')
  elif n2 > 1:
    K1 = 1.0*self.getK(population2, sndClass1)
    K2 = 1.0*self.getK(population2, sndClass2)
    K = K1+K2
    if K > 0.0:
      D1 = 1.0*self.getL(population1, population2, sndClass1)
      D2 = 1.0*self.getL(population1, population2, sndClass2)
      D = D1+D2
      a = b = 0.0
      for i in range(1, n2):
        a += 1.0/i
```

```
            b += 1.0/(i*i)
        T = D/(K/a) - 1.0
        theta1 = (K1+D1)/(a + 1.0 + T)
        theta2 = (K2+D2)/(a + 1.0 + T)
        EK1 = a*theta1
        EK2 = a*theta2
        ED1 = (1.0+T)*theta1
        ED2 = (1.0+T)*theta2
        VK1 = EK1 + b*theta1**2
        VK2 = EK2 + b*theta2**2
        VD1 = ED1 + theta1**2
        VD2 = ED2 + theta2**2
        X2 = 0.0
        if VK1 != 0.0: X2 += (K1-EK1)**2/VK1
        if VK2 != 0.0: X2 += (K2-EK2)**2/VK2
        if VD1 != 0.0: X2 += (D1-ED1)**2/VD1
        if VD2 != 0.0: X2 += (D2-ED2)**2/VD2
        p = chiSquared.pValue(X2, 1)
      else:
        p = float('nan')
    else:
      p = float('nan')
    return p


# D_FL_SimpleIndel is Fu and Li's D for simple indels.
# The first population is the ingroup, and the second is the outgroup.
# Simple indels in the ingroup that can't be polarized using the outgroup are ignored.
Def getD_FL_SimpleIndel(self, population1, population2, indelClass = '_all'):
    n = self.sample.taxonCounts[population1]
    eta = 0
    eta_e = 0
    for indel in self.sample.simpleIndels:
      if indelClass in indel.indelClasses:
        freqs1 = indel.alleleFrequencies[population1]
        freqs2 = indel.alleleFrequencies[population2]
        if len(freqs1) > 1 and len(freqs2) == 1 and freqs2.keys()[0] in freqs1.keys():
          eta += 1
          if freqs1[freqs2.keys()[0]] == n-1: eta_e += 1
    if n > 1 and eta > 0:
      a = 0.0
      b = 0.0
      for i in range(1, n):
        a += 1.0/i
        b += 1.0/(i*i)
      if n == 2:
        c = 1.0
      else:
        c = 2*(n*a - 2*(n-1))/((n-1)*(n-2))
      v = 1 + a*a*(c - (n+1)/(1.0*(n-1)))/(b + a*a)
      u = a-1-v
      D_FL = (eta - a*eta_e)/math.sqrt(u*eta + v*eta*eta)
    else:
      D_FL = 0.0
```

```
    return D_FL
  # F_FL_SimpleIndel is Fu and Li's F for simple indels.
  # The first population is the ingroup, and the second is the outgroup.
  # Simple indels in the ingroup that can't be polarized using the outgroup are ignored.
  Def getF_FL_SimpleIndel(self, population1, population2, indelClass = '_all'):
    n = self.sample.taxonCounts[population1]
    pi = 0
    eta = 0
    eta_e = 0
    for indel in self.sample.simpleIndels:
      if indelClass in indel.indelClasses:
        freqs1 = indel.alleleFrequencies[population1]
        freqs2 = indel.alleleFrequencies[population2]
        if len(freqs1) > 1 and len(freqs2) == 1 and freqs2.keys()[0] in freqs1.keys():
          for al1 in freqs1.keys():
            for al2 in freqs1.keys():
              if al1 != al2: pi += freqs1[al1]*freqs1[al2]
          eta += 1
          if freqs1[freqs2.keys()[0]] == n-1: eta_e += 1
    if n > 1 and eta > 0:
      pi /= 1.0*n*(n-1)
      a = 0.0
      b = 0.0
      for i in range(1, n):
        a += 1.0/i
        b += 1.0/(i*i)
      if n == 2:
        c = 1.0
      else:
        c = 2*(n*a – 2*(n-1))/((n-1)*(n-2))
      v = (c + 2.0*(n*n + n + 3)/(9*n*(n-1)) – 2.0/(n-1))/(b + a*a)
      u = (1 + (n+1)/(3.0*(n-1)) – 4.0*(n+1)*(a + 1.0/n – 2.0*n/(n+1))/((n-1)**2))/a – v
      F_FL = (pi – eta_e)/math.sqrt(u*eta + v*eta*eta)
    else:
      F_FL = 0.0
    return F_FL_SimpleIndel
  # thetaH_SimpleIndel is Fay and Wu's estimator of theta (4 N_e u) for simple indels.
  # The first population is the ingroup, and the second is the outgroup.
  # Simple indels in the ingroup that can't be polarized using the outgroup are ignored.
  Def get_thetaH_SimpleIndel(self, population1, population2, indelClass = '_all'):
    if self.cache.has_key(('thetaH_SimpleIndel', population1, population2, indelClass)):
      return self.cache[('thetaH_SimpleIndel', population1, population2, indelClass)]
    else:
      thetaH = 0.0
      n = self.sample.taxonCounts[population1]
      if n > 1:
        S = {}
        for i in range(1, n): S[i] = 0
        for indel in self.sample.simpleIndels:
          if indelClass in indel.indelClasses:
            freqs1 = indel.alleleFrequencies[population1]
            freqs2 = indel.alleleFrequencies[population2]
            if len(freqs1) > 1 and len(freqs2) == 1 and freqs2.keys()[0] in freqs1.keys(): S[n-
freqs1[freqs2.keys()[0]]] += 1
```

```
            for i in range(1, n): thetaH += S[i]*i*i
            thetaH *= 2.0/(n*(n-1))
        self.cache[('thetaH_SimpleIndel', population1, population2, indelClass)] = thetaH
        return thetaH
  def get_thetaH_SimpleIndelPerSite(self, population1, population2, indelClass = '_all', siteClass =
'_all'):
    length = self.sample.unambiguousSiteCounts[siteClass]
    if length > 0:
      thetaH_PerSite = self.get_thetaH_SimpleIndel(population1, population2, indelClass)/length
    else:
      thetaH_PerSite = 0.0
    return thetaH_PerSite
  # H_SimpleIndel is Fay and Wu's H for simple indels.
  # The first population is the ingroup, and the second is the outgroup.
  # Simple indels in the ingroup that can't be polarized using the outgroup are ignored.
  Def getH_SimpleIndel(self, population1, population2, indelClass = '_all'):
    n = self.sample.taxonCounts[population1]
    theta_pi = 0
    if n > 1:
      S = {}
      for i in range(1, n): S[i] = 0
      for indel in self.sample.simpleIndels:
        if indelClass in indel.indelClasses:
          freqs1 = indel.alleleFrequencies[population1]
          freqs2 = indel.alleleFrequencies[population2]
          if len(freqs1) > 1 and len(freqs2) == 1 and freqs2.keys()[0] in freqs1.keys(): S[n-
freqs1[freqs2.keys()[0]]] += 1
      for i in range(1, n): theta_pi += S[i]*i*(n-i)
      theta_pi *= 2.0/(n*(n-1))
    H = theta_pi-self.get_thetaH_SimpleIndel(population1, population2, indelClass)
    return H
  def getH_SimpleIndelPerSite(self, population1, population2, indelClass = '_all', siteClass = '_all'):
    length = self.sample.unambiguousSiteCounts[siteClass]
    if length > 0:
      H_PerSite = self.getH_SimpleIndel(population1, population2, indelClass)/length
    else:
      H_PerSite = 0.0
    return H_PerSite

class Slider:
  def __init__(self, alignment, halfWidth, slide, populations = {}, siteClasses = {}, options = {}):
    self.halfWidth = halfWidth
    self.slide = slide
    self.statisticsHash = {}
    length = alignment.getSiteCount()
    center = 0
    while center < length:
      windowAlignment = MyAlignment()
      left = max(center-halfWidth, 0)
      right = min(center+halfWidth+1, length)
      for taxon in range(alignment.getTaxonCount()):
windowAlignment.addSequence(alignment.getSequence(taxon)[left:right])
      windowSiteClasses = {}
      for siteClass in siteClasses:
```

```
            windowSiteClasses[siteClass] = []
            for site in siteClasses[siteClass]:
               if left <= site and site < right: windowSiteClasses[siteClass].append(site-left)
         self.statisticsHash[center] = Statistics(windowAlignment, populations, windowSiteClasses, options)
         center += slide
   def get_thetaW_PerSite(self, population = '_all', sndClass = '_all', siteClass = '_all'):
      pairs = [(center, self.statisticsHash[center].get_thetaW_PerSite(population, sndClass, siteClass)) for
center in self.statisticsHash.keys()]
      pairs.sort()
      return pairs
   def get_piPerSite(self, population = '_all', sndClass = '_all', siteClass = '_all'):
      pairs = [(center, self.statisticsHash[center].get_piPerSite(population, sndClass, siteClass)) for center in
self.statisticsHash.keys()]
      pairs.sort()
      return pairs
   def getD_T(self, population = '_all', sndClass = '_all'):
      pairs = [(center, self.statisticsHash[center].getD_T(population, sndClass)) for center in
self.statisticsHash.keys()]
      pairs.sort()
      return pairs
   def get_thetaW_SimpleIndelPerSite(self, population = '_all', indelClass = '_all', siteClass = '_all'):
      pairs = [(center, self.statisticsHash[center].get_thetaW_SimpleIndelPerSite(population, indelClass,
siteClass)) for center in self.statisticsHash.keys()]
      pairs.sort()
      return pairs
   def get_piSimpleIndelPerSite(self, population = '_all', indelClass = '_all', siteClass = '_all'):
      pairs = [(center, self.statisticsHash[center].get_piSimpleIndelPerSite(population, indelClass,
siteClass)) for center in self.statisticsHash.keys()]
      pairs.sort()
      return pairs
   def getD_T_SimpleIndel(self, population = '_all', indelClass = '_all'):
      pairs = [(center, self.statisticsHash[center].getD_T_SimpleIndel(population, indelClass)) for center in
self.statisticsHash.keys()]
      pairs.sort()
      return pairs
   def get_piBW_PerSite(self, population = '_all', indelClass = '_all', siteClass = '_all'):
      pairs = [(center, self.statisticsHash[center].get_piBW_PerSite(population, indelClass, siteClass)) for
center in self.statisticsHash.keys()]
      pairs.sort()
      return pairs
   def get_rhoPerSite(self, population1, population2, sndClass = '_all', siteClass = '_all'):
      pairs = [(center, self.statisticsHash[center].get_rhoPerSite(population1, population2, sndClass,
siteClass)) for center in self.statisticsHash.keys()]
      pairs.sort()
      return pairs
   def getD_FL(self, population1, population2, sndClass = '_all'):
      pairs = [(center, self.statisticsHash[center].getD_FL(population1, population2, sndClass)) for center in
self.statisticsHash.keys()]
      pairs.sort()
      return pairs
   def getF_FL(self, population1, population2, sndClass = '_all'):
      pairs = [(center, self.statisticsHash[center].getF_FL(population1, population2, sndClass)) for center in
self.statisticsHash.keys()]
      pairs.sort()
```

```
        return pairs
    def get_thetaH_PerSite(self, population1, population2, sndClass = '_all', siteClass = '_all'):
        pairs = [(center, self.statisticsHash[center].get_thetaH_PerSite(population1, population2, sndClass,
siteClass)) for center in self.statisticsHash.keys()]
        pairs.sort()
        return pairs
    def getH_PerSite(self, population1, population2, sndClass = '_all', siteClass = '_all'):
        pairs = [(center, self.statisticsHash[center].getH_PerSite(population1, population2, sndClass,
siteClass)) for center in self.statisticsHash.keys()]
        pairs.sort()
        return pairs
    def getF_ST(self, population, sndClass = '_all'):
        pairs = [(center, self.statisticsHash[center].getF_ST(population, sndClass)) for center in
self.statisticsHash.keys()]
        pairs.sort()
        return pairs
    def getPairwiseF_ST(self, population1, population2, sndClass = '_all'):
        pairs = [(center, self.statisticsHash[center].getPairwiseF_ST(population1, population2, sndClass)) for
center in self.statisticsHash.keys()]
        pairs.sort()
        return pairs
    def get_rhoSimpleIndelPerSite(self, population1, population2, indelClass = '_all', siteClass = '_all'):
        pairs = [(center, self.statisticsHash[center].get_rhoSimpleIndelPerSite(population1, population2,
indelClass, siteClass)) for center in self.statisticsHash.keys()]
        pairs.sort()
        return pairs
    def getD_FL_SimpleIndel(self, population1, population2, indelClass = '_all'):
        pairs = [(center, self.statisticsHash[center].getD_FL_SimpleIndel(population1, population2,
indelClass)) for center in self.statisticsHash.keys()]
        pairs.sort()
        return pairs
    def getF_FL_SimpleIndel(self, population1, population2, indelClass = '_all'):
        pairs = [(center, self.statisticsHash[center].getF_FL_SimpleIndel(population1, population2,
indelClass)) for center in self.statisticsHash.keys()]
        pairs.sort()
        return pairs
    def get_thetaH_SimpleIndelPerSite(self, population1, population2, indelClass = '_all', siteClass =
'_all'):
        pairs = [(center, self.statisticsHash[center].get_thetaH_SimpleIndelPerSite(population1, population2,
indelClass, siteClass)) for center in self.statisticsHash.keys()]
        pairs.sort()
        return pairs
    def getH_SimpleIndelPerSite(self, population1, population2, indelClass = '_all', siteClass = '_all'):
        pairs = [(center, self.statisticsHash[center].getH_SimpleIndelPerSite(population1, population2,
indelClass, siteClass)) for center in self.statisticsHash.keys()]
        pairs.sort()
        return pairs

# A hap file is an input file to Hudson's program exhap (Hudson, 2001, Genetics 159:1805–1817;
http://home.uchicago.edu/~rhudson1).
# The first population is the ingroup, and the second is the outgroup.
Def AlignmentToHapFile(alignment,    ilename, populations, population1 = '_all', population2 = None):
    sample = Sample(alignment, populations)
    alleleLists = {}
```

```
ancestralAlleles = {}
for snd in sample.snds:
   freqs1 = snd.nucleotideFrequencies[population1]
   if len(freqs1) == 2:
      site = snd.site
      alleleLists[site] = snd.nucleotides[population1]
      ancestralAlleles[site] = '?'
      if population2 != None:
         freqs2 = snd.nucleotideFrequencies[population2]
         if len(freqs2) == 1 and freqs2.keys()[0] in freqs1.keys(): ancestralAlleles[site] = freqs2.keys()[0]
for indel in sample.simpleIndels+sample.complexIndels:
   freqs1 = indel.alleleFrequencies[population1]
   if len(freqs1) == 2:
      site = indel.site
      alleleLists[site] = indel.alleles[population1]
      ancestralAlleles[site] = '?'
      if population2 != None:
         freqs2 = indel.alleleFrequencies[population2]
         if len(freqs2) == 1 and freqs2.keys()[0] in freqs1.keys(): ancestralAlleles[site] = freqs2.keys()[0]
sites = alleleLists.keys()
sites.sort()
siteCount = len(sites)
taxonCount = sample.taxonCounts[population1]
handle = open(   ilename, 'w')
handle.write("%d %d\n" % (taxonCount, siteCount))
for site in sites: handle.write("%d " % site)
handle.write('\n')
handle.write("a ")
for site in sites: handle.write("%s " % ancestralAlleles[site])
handle.write('\n')
for taxon in range(taxonCount):
   handle.write("%d " % taxon)
   for site in sites: handle.write("%s " % alleleLists[site][taxon])
   handle.write('\n')
handle.close()
```

Generating statistical values

Our analyses was performed with three different loci with alleles sequenced from a natural *X. l. laevis* (XLL) population and several closely related *Xenopus* species (e.g. *X. amieti*: XA). In order to run these analyses you must first generate a FASTA or PHYLP (non-interleaved) alignment. The range of population sequences in the alignment is specified by the range "XLL" (27) and the location of the outgroup [27] . After some basic analyses, a sliding-window analyses using a window of 101 (50+1+50) bases and a slide of 10 bases is conducted with several of the summary statistics. Finally, a .hap file is generated with our data, which is useful if you have Richard Hudson's programs exhap and maxhap for estimating recombination (http://home.uchicago.edu/~rhudson1).

```
#! /usr/bin/env python

from populationGenetics import *

alignment = FASTA_FileToMyAlignment("GAD67.promoter.TBAa.txt")
populations = {'XLL': range(27), 'XA': [27]}
statistics = Statistics(alignment, populations)
print "XLL sequences: %d" % statistics.getTaxonCount('XLL')
print "XA sequences: %d" % statistics.getTaxonCount('XA')
print "total length: %d" % statistics.getSiteCount()
print "unambiguous length: %d" % statistics.getUnambiguousSiteCount()
print "informative length: %d" % statistics.getInformativeSiteCount()
print "K: %d" % statistics.getK('XLL')
print "singleton count: %d" % statistics.getSingletonCount('XLL')
print "thetaW per site: %g" % statistics.get_thetaW_PerSite('XLL')
print "pi per site: %g" % statistics.get_piPerSite('XLL')
print "D_T: %g" % statistics.getD_T('XLL')
print "L: %d" % statistics.getL('XLL', 'XA')
print "rho per site: %g" % statistics.get_rhoPerSite('XLL', 'XA')
print "D_FL: %g" % statistics.getD_FL('XLL', 'XA')
print "H: %g" % statistics.getH('XLL', 'XA')
print "thetaH_Persite: %g" % statistics.get_thetaH('XLL', 'XA')
print "pi: %g" % statistics.get_pi('XLL')
print "K_SimpleIndel: %g" % statistics.getK_SimpleIndel('XLL')
print "thetaW_SimpleIndel: %g" % statistics.get_thetaW_SimpleIndel('XLL')
print "piSimpleIndel: %g" % statistics.get_piSimpleIndel('XLL')
print "piSimpleIndelPerSite: %g" % statistics.get_piSimpleIndelPerSite('XLL')
```

```
print "D_T_SimpleIndel: %g" % statistics.getD_T_SimpleIndel('XLL')
print "D_FL_SimpleIndel: %g" % statistics.getD_FL_SimpleIndel('XLL','XA')
print "piBW: %g" % statistics.get_piBW('XLL')
print "piBW_PerSite: %g" % statistics.get_piBW_PerSite('XLL')
slider = Slider(alignment, 50, 10, populations)
D_Ts = slider.getD_T('XLL')
file = open("GAD67promoter_XA_D_T", 'w')
for D_T in D_Ts: file.write("%d %g\n"%D_T)
file.close()
D_FLs = slider.getD_FL('XLL', 'XA')
file = open("GAD67promoter_XA_D_FL", 'w')
for D_FL in D_FLs: file.write("%d %g\n"%D_FL)
file.close()
H_PerSites = slider.getH_PerSite('XLL', 'XA')
file = open("GAD67promoter_XA_H_PerSite", 'w')
for H_PerSite in H_PerSites: file.write("%d %g\n"%H_PerSite)
file.close()
pi_PerSites = slider.get_piPerSite('XLL')
file = open("GAD67promoter_XA_pi_PerSite", 'w')
for pi_PerSite in pi_PerSites: file.write("%d %g\n"%pi_PerSite)
file.close()
piBW_PerSites = slider.get_piBW_PerSite('XLL')
file = open("GAD67promoter_XA_piBW_PerSite", 'w')
for piBW_PerSite in piBW_PerSites: file.write("%d %g\n"%piBW_PerSite)
file.close()
D_T_SimpleIndels= slider.getD_T_SimpleIndel('XLL')
file = open("GAD67promoter_XA_D_T_SimpleIndel", 'w')
for D_T_SimpleIndel in D_T_SimpleIndels: file.write("%d %g\n"%D_T_SimpleIndel)
file.close()
D_FL_SimpleIndels = slider.getD_FL_SimpleIndel('XLL', 'XA')
file = open("GAD67promoter_XA_D_FL_SimpleIndel", 'w')
for D_FL_SimpleIndel in D_FL_SimpleIndels: file.write("%d %g\n"%D_FL_SimpleIndel)
file.close()
AlignmentToHapFile(alignment, "GAD67.promoter_XA.hap", populations, 'XLL', 'XA')
```

Source code for running the HKA test

Summary code for implementing the HKA test of a neutral model is provided below. This program requires two different FASTA alignment files: One alignment of a region of evolutionary interests (i.e. promoter) and is defined as bindingAlignment, and a second alignment that will serve as a neutral proxy (i.e. intron) and is defined as nonbindingAlignment. A chi-squared distribution is implement in C and not provided here.

```
#! /usr/bin/env python

from populationGenetics import *

bindingAlignment = FASTA_FileToMyAlignment("GAD67.promoter.HKA.txt")
bindingLength = bindingAlignment.getSiteCount()
nonbindingAlignment = FASTA_FileToMyAlignment("GAD67.Intron15.CLUSTALa.txt")
nonbindingLength = nonbindingAlignment.getSiteCount()
alignment = MyAlignment.concatenation(bindingAlignment, nonbindingAlignment)
populations = {'XLL': range(24), 'XA': [24]}
siteClasses = {'binding': range(bindingLength), 'nonbinding': range(bindingLength,
bindingLength+nonbindingLength)}
statistics = Statistics(alignment, populations, siteClasses)
print "HKA p-value: %g" % statistics.getHKApValue('XLL', 'XA', 'binding', 'nonbinding')
```

Estimating ρ with Hudson's exhap and maxhap programs

To estimate ρ from sequence alignment data, we used R.H. software to generate a haplotype file (e.g. filename.hap) , and ran the following command in an UNIX shell

```
% exhap < YourData.hap > YourData.pairs
```

to convert a haplotype data file into a file with the necessary Hudson pairs data format. The command

```
% maxhap 1 h27rho 0.01 100 0.01 0 0 1 0 < YourData.pairs
```

is instructs maxhap to estimate ρ for a dataset. The command line arguments tell maxhap to use one two-site configuration file of two-locus sampling probabilities for 27 samples. Maxhap will begin to search for the maximum composite-likelihood score starting from 0.01 x 0.01 to a maximum value 100 x 0.01 in equally spaced increments. The last four values specify gene conversion parameters which were not included in our analysis.

Generating sample populations with Hudson's ms by coalescent simulation

Hudson's ms program is available online (http://home.uchicago.edu/~rhudson1/). To run the program download the ms tarball and cd to the downloaded directory in a UNIX shell. The command

```
% ms 27 100000 –s 46 –r 0 973 | sample_stats > SummaryStats_sample_stats
```

uses ms to generate random genealogies for 100,000 replicate sample populations of 27 haploid individuals at a 973 bp locus. A total of 46 mutations will be Poisson distributed along each replicate sample genealogy. The output data is then submitted to sample_stats to extract summary statistics for each replicate sample, which is imported into the file SummaryStats_sample_stats. To assess the statistical significance of an observed test statistic run

```
% cut –f 6 SummaryStats_sample_stats | ./stats 0.024
```

which uses the UNIX command cut to extract the 6^{th} column of data from the SummaryStats_sample_stats file and assess the value at the 2.4 percentile.

TBA evolutionary tree

In order to run a TBA alignment, a binary tree must be specified in a modified Newick format that details the evolutionary relationship of the sequence included. Branch lengths are not required. A tree for 27 *X l. laevis* alleles ('XLL') and four outgroups species ('XLS', 'XA', 'XC'. 'XM') is provided. The 'XLL1 XLL2' group merely seeds the alignment.

Example:

(((((((((((((((((((((((((((XLL1 XLL2) XLL3) XLL4) XLL5) XLL6 XLL7) XLL8) XLL9) XLL10) XLL11) XLL12) XLL13) XLL14) XLL15) XLL16) XLL17) XLL18) XLL19) XLL20) XLL21) XLL22) XLL23) XLL24) XLL25) XLL26) XLL27) XLS) XA) XC) XM)

ClustalX 2.0.11 alignment of the upstream region of *xGAD67.*

Sequences for 27 *X. l. laevis* alleles (XLL1 – XLL26) and four closely related outgroup species (XA, *X. amieti*; XC, *X. clivii*; XLS, *X. l. sudanensis*; XM, *X. muelleri*) were aligned under default parameters: gap opening and extension penalties 15 and 6.66, respectively. The transcriptional start site is indicated by +1 and the initiating methionine is underlined. Sequences proceed from 5' to 3' and include a portion of the first exon. Conserved sites are indicated by an asterisk (*).

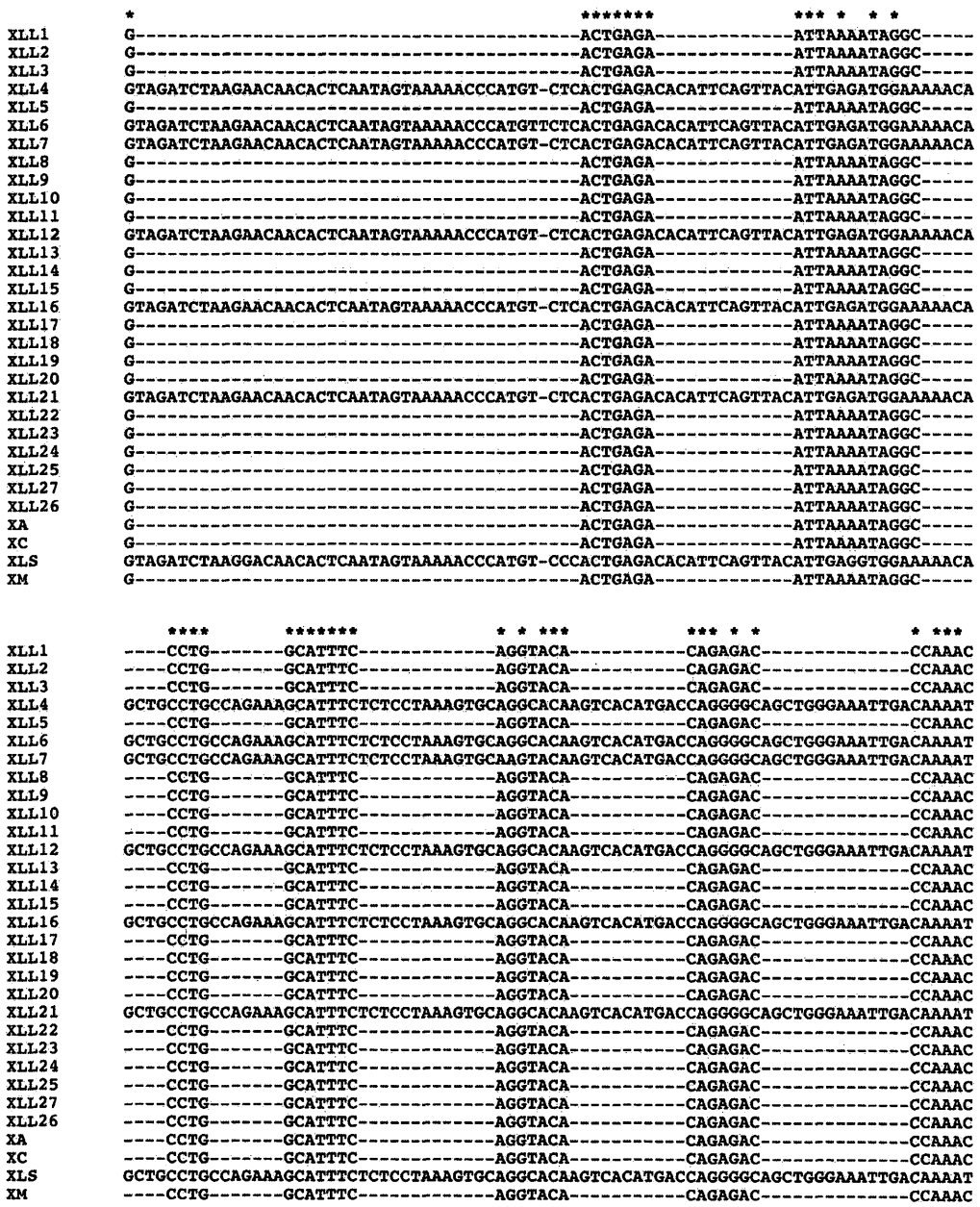
```
          ****************** ********************************** *************************
XLL1      ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL2      ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL3      ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL4      ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL5      --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL6      ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL7      ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL8      ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL9      ACACCAGCACGTTCTCCATCCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL10     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL11     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL12     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL13     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL14     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL15     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL16     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL17     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL18     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL19     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL20     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL21     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL22     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL23     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL24     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL25     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL27     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL26     ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XA        ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XC        ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLS       ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XM        ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC

          ******* ********************************** ********* ****************************
XLL1      AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL2      AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL3      AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL4      AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL5      AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL6      AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL7      AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL8      AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL9      AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL10     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL11     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL12     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL13     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL14     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL15     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL16     AAAGCACGAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL17     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL18     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL19     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL20     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL21     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL22     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL23     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL24     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL25     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL27     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL26     AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XA        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XC        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCCCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLS       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XM        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTGGAAATAGCACTAATGTCGGTATAAACAT
```

```
          ********************* ** ***** * * ******** ******  ***** ************** ********
XLL1      TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL2      TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL3      TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGAACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT
XLL4      TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTTTTCGCAGTCACATTTCGAAGCTT
XLL5      TAATGGAGAAGGATTTTGTAAGAGAAAGGAATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL6      TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL7      TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL8      TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL9      TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL10     TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL11     TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL12     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL13     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL14     TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL15     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL16     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL17     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL18     TAATGGAGAAGGATTTTGTAAGAAAAAGGGACTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL19     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL20     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL21     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATT-CATCCTTGAACAGACATGTATTGTTCGCAGTCACATTCCGAAGCTT
XLL22     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL23     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL24     TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL25     TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGAACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT
XLL27     TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL26     TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XA        TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGGACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT
XC        TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLS       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XM        TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGGACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT

          ********************** * ******************** ******** ******* ** **************
XLL1      GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL2      GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL3      GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL4      GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL5      GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL6      GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL7      GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL8      GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTGGTTCTGCCTGCTGCTCCTATAAATT---
XLL9      GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL10     GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL11     GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL12     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL13     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL14     GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL15     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL16     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL17     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL18     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL19     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL20     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL21     GAAGGTCAAAGTGTCTCACTAGAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL22     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL23     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL24     GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL25     GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL27     GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCTTGCTGCTCCTATAAATT---
XLL26     GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XA        GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XC        GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLS       GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGTTGCTCCTATAAATTTCA
XM        GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
```

```
          * ******************************** ************** *** ************ ***** *** *****
XLL1      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL2      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL3      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL4      CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTTTACCTATGTG
XLL5      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL6      CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL7      CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL8      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL9      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL10     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL11     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL12     CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL13     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL14     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL15     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL16     CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL17     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL18     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCCATGTG
XLL19     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL20     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL21     CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL22     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL23     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL24     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL25     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCCCTAAAACAAAAA-TACTT-ACCTATGTG
XLL27     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTGCCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLL26     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XA        CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XC        CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XLS       CCATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTT-ACCTATGTG
XM        CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAAATACTT-ACCTATGTG

           ************************  *************** ** ** ************ * *    ***
XLL1      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL2      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL3      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL4      CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTTGGGG-AAAGATACTGTATGTTAAAGGGATACTGTCATGGGAA
XLL5      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL6      CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTTGGGG-AAAGATACTGTATGTTAAAGGGATACTGTCATGGGAA
XLL7      CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTTGGGG-AAAGATACTGTATGTTAAAGGGATACTGTCATGGGAA
XLL8      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL9      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAGGAGACTGTATGTTAAGGTGACTGTGT--------
XLL10     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL11     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL12     CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTTGGGG-AAAGATACTGTATGTTAAAGGGATACTGTCATGGGAA
XLL13     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL14     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL15     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL16     CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTTGGGG-AAAGATACTGTATGTTAAAGGGATACTGTCATGGTAA
XLL17     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL18     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL19     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL20     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL21     CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTTGGGG-AAAGATACTGTATGTTAAAGGGATACTGTCATGGGAA
XLL22     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL23     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL24     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL25     CTGGCATTTTGGACATTTGTAGACACGACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL27     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XLL26     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGGGAAAGAGACTGTATGTTAAGGTGACTGTGT--------
XA        CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
XC        CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGCCTGTGT--------
XLS       CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTTGGGG-AAAGATACTGTATGTTAAAGGGATACTGTCATGGGA-
XM        TTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGG-AAAGAGACTGTATGTTAAGGTGACTGTGT--------
```

```
                                         *** **                           ***  * *
XLL1      ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL2      ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL3      ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL4      GAAAAAAATATTTTCAAAATGAATCAGTTAATAGTGCTGCTCCAGCAGAATTCTGCACTGAAATCCATTTCTCAAAAGAG
XLL5      ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL6      GAAAAAAATATTTTCAAAATGAATCAGTTAATAGTGCTGCTCCAGCAGAATTCTGCACTGAAATCCATTTCTCAAAAGAG
XLL7      GAAAAAAATATTTTCAAAATGAATCAGTTAATAGTGCTGCTCCAGCAGAATTCTGCACTGAAATCCATTTCTCAAAAGAG
XLL8      ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL9      ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL10     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL11     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL12     GAAAAAAATATTTTCAAAATGAATCAGTTAATAGTGCTGCTCCAGCAGAATTCTGCACTGAAATCCATTTCTCAAAAGAG
XLL13     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL14     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL15     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL16     GAAAAAAATATTTTCAAAATGAATCAGTTAATAGTGCTGCTCCAGCAGAATTCTGCACTGAAATCCATTTCTCAAAAGAG
XLL17     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL18     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL19     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL20     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL21     GAAAAAAATATTTTCAAAATGAATCAGTTAATAGTGCTGCTCCAGCAGAATTCTGCACTGAAATCCATTTCTCAAAAGAG
XLL22     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL23     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL24     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL25     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL27     ------------------------------ATACTG---------------------------TCCCTTCC---------
XLL26     ------------------------------ATACTG---------------------------TCCCTTCC---------
XA        ------------------------------ATACTG---------------------------TCCCTTCC---------
XC        ------------------------------ATACTG---------------------------TCCCCTCC---------
XLS       -AAAAAATTTTTTTCAAAATGCATCAGTTAATAGTGCTGCTCCAGCAGAATTCTGCACTGAAATCCATTTC---------
XM        ------------------------------ATACTG---------------------------TCCCTTCC---------

                                                     *   ******             **
XLL1      -------------------------------------------CTGAACATTA-------------GC------------
XLL2      -------------------------------------------CTGAACATTA-------------GC------------
XLL3      -------------------------------------------CTGAACATTA-------------GC------------
XLL4      CAAACAGATTTTTTTATATTCAATTTTGAAATCTGACATGGGGCTAGACATTATGTCAATTTCCCAGCTACCCCTAGTCA
XLL5      -------------------------------------------CTGAACATTA-------------GC------------
XLL6      CAAACAGATTTTTTTATATTCAATTTTGAAATCTGACATGGGGCTAGACATTATGTCAATTTCCCAGCTACCCCTAGTCA
XLL7      CAAACAGATTTTTTTATATTCAATTTTGAAATCTGACATGGGGCTAGACATTATGTCAATTTCCCAGCTACCCCTAGTCA
XLL8      -------------------------------------------CTGAACATTA-------------GC------------
XLL9      -------------------------------------------CCGAACATTA-------------GC------------
XLL10     -------------------------------------------CTGAACATTA-------------GC------------
XLL11     -------------------------------------------CTGAACATTA-------------GC------------
XLL12     CAAACAGATTTTTTTATATTCAATTTTGAAATCTGACATGGGGCTAGACATTATGTCAATTTCCCAGCTACCCCTAGTCA
XLL13     -------------------------------------------CTGAACATTA-------------GC------------
XLL14     -------------------------------------------CTGAACATTA-------------GC------------
XLL15     -------------------------------------------CTGAACATTA-------------GC------------
XLL16     CAAACAGATTTTTTTATATTCAATTTTGAAATCTGACATGGGGCTAGACATTATGTCAATTTCCCAGCTACCCCTAGTCA
XLL17     -------------------------------------------CTGAACATTA-------------GC------------
XLL18     -------------------------------------------CTGAACATTA-------------GC------------
XLL19     -------------------------------------------CTGAACATTA-------------GC------------
XLL20     -------------------------------------------CTGAACATTA-------------GC------------
XLL21     CAAACAGATTTTTTTATATTCAATTTTGAAATCTGACATGGGGCTAGACATTATGTCAATTTCCCAGCTACCCCTAGTCA
XLL22     -------------------------------------------CTGAACATTA-------------GC------------
XLL23     -------------------------------------------CTGAACATTA-------------GC------------
XLL24     -------------------------------------------CTGAACATTA-------------GC------------
XLL25     -------------------------------------------CTGAACATTA-------------GC------------
XLL27     -------------------------------------------CTGAACATTA-------------GC------------
XLL26     -------------------------------------------CTGAACATTA-------------GC------------
XA        -------------------------------------------CTGAACATTA-------------GC------------
XC        -------------------------------------------CTGAACATTA-------------GC------------
XLS       -----AGATTTTTT-------------------TGACATGGGGCTAGACATTATGTCAATTTCCCAGCTGCCCTTCGTCA
XM        -------------------------------------------CTGAACATTA-------------GC------------
```

```
XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    TGTGACTTGTGCTCTGATAAACTTCAATCACTCTTTACTGCTGTACTGCAAGTTGGAGTGATATCACCCACCTCCCTCCC
XLL5    ----------------------------------------------------------------------------------------
XLL6    TGTGACTTGTGCTCTGATAAACTTCAATCACTCTTTACTGCTGTACTGCAAGTTGGAGTGATATCACCCACCTCCCTCCC
XLL7    TGTGACTTGTGCTCTGATAAACTTCAGTCACTCTTTACTGCTGTACTGCAAGTTGGAGTGATATCACCCACCTCCCTCCC
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   ----------------------------------------------------------------------------------------
XLL12   TGTGACTTGTGCTCTGATAAACTTCAATCACTCTTTACTGCTGTACTGCAAGTTGGAGTGATATCACCCACCTCCCTCCC
XLL13   ----------------------------------------------------------------------------------------
XLL14   ----------------------------------------------------------------------------------------
XLL15   ----------------------------------------------------------------------------------------
XLL16   TGTGACTTGTGCTCTGATAAACTTCAATCACTCTTTACTGCTGTACTGCAAGTTGGAGTGATATCACCCACCTCCCTCCC
XLL17   ----------------------------------------------------------------------------------------
XLL18   ----------------------------------------------------------------------------------------
XLL19   ----------------------------------------------------------------------------------------
XLL20   ----------------------------------------------------------------------------------------
XLL21   TGTGACTTGTGCTCTGATAAACTTCAATCACTCTTTACTGCTGTACTGCAAGTTGGAGTGATATCACCCACCTCCCTCCC
XLL22   ----------------------------------------------------------------------------------------
XLL23   ----------------------------------------------------------------------------------------
XLL24   ----------------------------------------------------------------------------------------
XLL25   ----------------------------------------------------------------------------------------
XLL27   ----------------------------------------------------------------------------------------
XLL26   ----------------------------------------------------------------------------------------
XA      ----------------------------------------------------------------------------------------
XC      ----------------------------------------------------------------------------------------
XLS     TGTGACTTATGCTCTGATAAACTTCAATCACTCTTTACTGCTGTACTGCAAGTTGGAGTGAAATCACCC-CCTCCCTCCC
XM      ----------------------------------------------------------------------------------------

                             ****  ***   *  * * *    ** * *  * *                      *** *
XLL1    -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL2    -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL3    -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCTG
XLL4    CCCCCCCC-AGCAGCCAAACAAAAGAACAATGGGAAGGTAACCAGATAGCAGCTCCCTAACACAAGATAACAGCTGCCTG
XLL5    -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL6    CCCCCCCC-AGCAGCCAAACAAAAGAACAATGGGAGGGTAACCAGATAGCAGCTCCCTAGCACAAGATAACAGCTGCCTG
XLL7    CCCCCCCCCCAGCAGCCAAACAAAAGAACAATGGGAAGGTAACCAGATAGCAGCTCCCTAACACAAGATAACAGCTGCCTG
XLL8    -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL9    -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL10   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL11   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL12   CCCCCCCC-AGCAGCCAAACAAAAGAACAATGGGAAGGTAACCAGATAGCAGCTCCCTAACACAAGATAACAGCTGCCTG
XLL13   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL14   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCGG-----------------------GCCCG
XLL15   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL16   CCCCCCCCCAGCAGCCAAACAAAAGAACAATGGGAAGGTAACCAGATAGCAGCTCCCTAACACAAGATAACAGCTGCCTG
XLL17   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL18   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL19   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL20   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL21   CCCCCCCC-AGCAGCCAAACAAAAGAACAATGGGAAGGTAACCAGATAGCAGCTCCCTAACACAAGATAACAGCTGCCTG
XLL22   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL23   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL24   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL25   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCTG
XLL27   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLL26   -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XA      -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCTG
XC      -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCCG
XLS     CCCCCCC--AGCAGCCAAACAACAGAACAATGGGAAGGTAACCAGATAACAGCTCCCTAACACAAGATAACAGCTGCCTG
XM      -----------------AACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAG-----------------------GCCTG
```


XLL1
XLL2
XLL3
XLL4
XLL5
XLL6
XLL7
XLL8
XLL9
XLL10
XLL11
XLL12
XLL13
XLL14
XLL15
XLL16
XLL17
XLL18
XLL19
XLL20
XLL21
XLL22
XLL23
XLL24
XLL25
XLL27
XLL26
XA
XC
XLS
XM

```
XLL1      AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLL2      AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLL3      AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCGCAATCAGC-
XLL4      GTCTAGCCCCATGTCAGATTTCAAATTTGAATATAAAAAAA-TGTGTTTGCTCTTTTGAGAAATGGATTTCAGTGCAGAA
XLL5      A-----------------------------------------------------------------------------
XLL6      GTCTAGCCCCATGTCAGATTTCAAATTTGAATATAAAAAAA-TGTGTTTGCTCTTTTGAGAAATGGATTTCAGTGCAGAA
XLL7      GTCTAGCCCCATGTCAGATTTCAAATTTGAATATAAAAAAA-TGTGTTTGCTCTTTTGAGAAATGGATTTCAGTGCAGAA
XLL8      A-----------------------------------------------------------------------------
XLL9      A-----------------------------------------------------------------------------
XLL10     A-----------------------------------------------------------------------------
XLL11     A-----------------------------------------------------------------------------
XLL12     GTCTAGCCCCATGTCAGATTTCAAATTTGAATATAAAAAAA-TGTGTTTGCTCTTTTGAGAAATGGATTTCAGTGCAGAA
XLL13     AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLL14     A-----------------------------------------------------------------------------
XLL15     AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLL16     GTCTAGCCCCATGTCCGATTTCAACTTTGAATATAAAAAAA-TGTCTTTGCTCTTTTGAGAAATGGATTTCAGTGCAGAA
XLL17     AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLL18     AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLL19     AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLL20     AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLL21     GTCTAGCCCCATGTCAGATTTCAAATTTGAATATAAAAAAA-TGTGTTTGCTCTTTTGAGAAATGGATTTCAGTGCAGAA
XLL22     AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLL23     AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLL24     A-----------------------------------------------------------------------------
XLL25     AC-------------AGGTCTGGGCTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCGCAATCAGC-
XLL27     AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLL26     A-----------------------------------------------------------------------------
XA        AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCGCAATCAGC-
XC        AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCCCAATCAGC-
XLS       GTCTAGCCCCATGTCAGATTTCAAATTTGAATATAAAAAAAATGTGTTTGCTCTTTTGAGAAATGGATTTCAGTGCAGAA
XM        AC-------------AGGTCTGGACTGAGAATTAAAATAGG---CCCTGGCATTTCAGGTACACAGAGGCGCAATCAGC-


XLL1      --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL2      --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL3      --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL4      GTCTGCTGGAGCAGCACTATTAACTGATTCATTTTGAAAAAAAAAAATTTT--CCCATGACAGTATCCCTTTAAGGTGACT
XLL5      -----------------------------------------------------------------------------
XLL6      GTCTGCTGGAGCAGCACTATTAACTGATTCATTTTGAAAAAAATTTTTTTTTTCCCATGACAGTATCCCTTTAAGGTGACT
XLL7      GTCTGCTGGAGCAGCACTATTAACTGATTCATTTTGAAAAAAAAATTTTT--CCCATGACAGTATCCCTTTAAGGTGACT
XLL8      -----------------------------------------------------------------------------
XLL9      -----------------------------------------------------------------------------
XLL10     -----------------------------------------------------------------------------
XLL11     -----------------------------------------------------------------------------
XLL12     GTCTGCTGGAGCAGCACTATTAACTGATTCATTTTGAAAAAAATTTTTTTTTCCCATGACAGTATCCCTTTAAGGTGACT
XLL13     --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL14     -----------------------------------------------------------------------------
XLL15     --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL16     GTC-GCTGGAGCAGCACTATTAACTGATTCATTTTGAAAAAAATTTTTTTTTCCCATGACAGTATCCCTTTAAGGTGACT
XLL17     --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL18     --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL19     --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL20     --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL21     GTCTGCTGGAGCAGCACTATTAACTGATTCATTTTGAAAAAAAAATTTTTTT-CCCATGACAGTATCCCTTTAAGGTGACT
XLL22     --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACCTTCTATGGCACCT
XLL23     --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL24     -----------------------------------------------------------------------------
XLL25     --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL27     --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLL26     -----------------------------------------------------------------------------
XA        --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XC        --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
XLS       GTCTGCTGGAGTAGCACTATTAACTGATTCATTTTGAAAAACATTTTTTT--CCCATGACAGTATCCCTTTAAGGTGACT
XM        --------------CCCCACCAGCCCACTAAAT---------------------------ACCGACTTTCTATGGCACCT
```

```
XLL1        ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL2        ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL3        ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL4        GTGTATACTGTCCCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTGGACTGAGAAT
XLL5        ------------------------------------------------------------------------------
XLL6        GTGTATACTGTCCCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTGGACTGAGAAT
XLL7        GTGTATACTGTCCCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTGGACTGAGAAT
XLL8        ------------------------------------------------------------------------------
XLL9        ------------------------------------------------------------------------------
XLL10       ------------------------------------------------------------------------------
XLL11       ------------------------------------------------------------------------------
XLL12       GTGTATACTGTCCCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTGGACTGAGAAT
XLL13       ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL14       ------------------------------------------------------------------------------
XLL15       ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL16       GTGTATACTGTCCCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTAGACTGAGAAT
XLL17       ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL18       ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL19       ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL20       ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL21       GTGTATACTGTCCCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGCCTGGACTGAGAAT
XLL22       ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL23       ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL24       ------------------------------------------------------------------------------
XLL25       ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL27       ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLL26       ------------------------------------------------------------------------------
XA          ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XC          ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------
XLS         GTGTATACTGTCCCTTTCCTGAGCATTAGCAATAGGAGATTTAGTGAAATAGGTCCTGGTATCAGGCTCGGACTGAGAAT
XM          ---TATAGCAGCCCCTCT---GGCATTTGCCAGAA-----------------------------------------------

                                                    ******** ******** ********************

XLL1        ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL2        ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL3        ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL4        TAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL5        ------------------------------------------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL6        TAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL7        TAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL8        ------------------------------------------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL9        ------------------------------------------GCCCCACCTGCCCACTAGATACTGACTTTCTATGGCAC
XLL10       ------------------------------------------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL11       ------------------------------------------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL12       TAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL13       ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL14       ------------------------------------------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL15       ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL16       TAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL17       ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL18       ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL19       ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL20       ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL21       TAAAATAGGCCCTGGCATTTCAGGTACACAGAGGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL22       ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL23       ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL24       ------------------------------------------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL25       ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL27       ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLL26       ------------------------------------------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XA          ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XC          ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
XLS         TAAAATAGGCCCTGGCATTTCAGTTACACAGAGGCCCAAACAGCCCCACCAGCCCACTAAATACTGACTTTCTATGGCAC
XM          ---------CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCAC
```

```
                  ****************** ************ ****************   * ************** ************
XLL1              CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL2              CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL3              CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL4              CTTATAGCAGCCCCTCTGCCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL5              CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL6              CTTATAGCAGCCCCTCTGCCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL7              CTTATAGCAGCCCCTCTGCCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL8              CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL9              CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL10             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCTGGGTCTGCCTGGTATTAATCATACATCCCTT
XLL11             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL12             CTTATAGCAGCCCCTCTGCCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL13             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL14             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL15             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL16             CTTATAGCAGCCCCTCTGCCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL17             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL18             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL19             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL20             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL21             CTTATAGCAGCCCCTCTGCCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL22             CTTATAGCAGCCCCTCTGGCATTTGCCAGAATCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL23             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL24             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL25             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAACCATACATCCCTT
XLL27             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLL26             CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCTGGGCCTGCCTGGTATTAATCATACATCCCTT
XA                CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XC                CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT
XLS               CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCAAGCCTGCCTGGTATTAATCATACATCCCTT
XM                CTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTT

                  ******** ********** ************************** ************************************
XLL1              CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL2              CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL3              CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL4              CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL5              CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL6              CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL7              CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL8              CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL9              CTCTATCTATTTGCTCAGTTAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL10             CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL11             CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL12             CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL13             CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL14             CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL15             CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL16             CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL17             CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL18             CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL19             CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL20             CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL21             CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL22             CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL23             CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL24             CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL25             CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL27             CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLL26             CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATT
XA                CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XC                CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XLS               CTCTATCTATTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
XM                CTCTATCTGTTTGCTCAGTCAACAGACTGTGCCTGTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATT
```

```
          ******* ********************************** ************************************* ****
XLL1      CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL2      CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL3      CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL4      CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL5      CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL6      CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL7      CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL8      CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL9      CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL10     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL11     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL12     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL13     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL14     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL15     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL16     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL17     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL18     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL19     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL20     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL21     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL22     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL23     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL24     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACGTTAG
XLL25     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL27     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLL26     CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XA        CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XC        CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XLS       CGCAGCATTTGATTAGTGGAGTGGAATCTGTGTGATCAAGTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG
XM        CGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTGATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAG

          *** ********** ***************** ************************************* * ********
XLL1      GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL2      GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL3      GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL4      GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL5      GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL6      GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL7      GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL8      GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCTCAAAGGCA
XLL9      GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL10     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL11     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL12     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL13     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL14     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL15     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL16     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL17     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL18     GTTTTTAACTTGCATGGCTGCTTTCCATTACCCGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACCCCCAAAGGCA
XLL19     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL20     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL21     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL22     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL23     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL24     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL25     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL27     GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLL26     GTTTTTAACTTGCACGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XA        GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XC        GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XLS       GTTCTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
XM        GTTTTTAACTTGCATGGCTGCTTTCCATTACCAGAGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCA
```

```
          **** ** ***************************** **********  ******* ******** *********** ***
XLL1      GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL2      GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL3      GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL4      GGCCCTGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATT
XLL5      GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL6      GGCCCTGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATT
XLL7      GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL8      GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGGGCTAAGATCTCCATCTGATCTATT
XLL9      GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL10     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL11     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL12     GGCCCTGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATT
XLL13     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCACTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL14     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGCATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL15     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL16     GGCCCTGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCCATT
XLL17     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL18     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL19     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL20     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL21     GGCCCTGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATT
XLL22     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL23     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL24     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL25     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL27     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLL26     GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XA        GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XC        GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XLS       GGCCATGCCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
XM        GGCCATGTCATATAAATAGTCTTTCCACTGGCCTCATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATT
                                                  +1

          ******************************************
XLL1      GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL2      GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL3      GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL4      GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL5      GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL6      GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL7      GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCA-
XLL8      GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL9      GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL10     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL11     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCA-
XLL12     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL13     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCA-
XLL14     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL15     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCA-
XLL16     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCA-
XLL17     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCA-
XLL18     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCA-
XLL19     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCT-
XLL20     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCT-
XLL21     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL22     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL23     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL24     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL25     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL27     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLL26     GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCA-
XA        GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XC        GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XLS       GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
XM        GCTTGATAGCCTGACCATGCTCAGGAAAGAGAACACCCTGCCAA
```

TBA alignment of the upstream region of *xGAD67*.

Twenty seven X. l. laevis alleles (XLL1 – XLL26) and four closely related outgroup species (XA, *X. amieti*; XC, *X. clivii*; XLS, *X. l. sudanensis*; XM, *X. muelleri*) were aligned using TBA with a modified Newick format parameter file (provided in the Appendix) describing the evolutionary relationship of outgroup species (Evans et al., 2004). The transcriptional start site is indicated by +1 and the initiating methionine is underlined. Sequences proceed from 5' to 3' and include a portion of the first exon. Conserved sites are indicated by an asterisk (*).

```
                ********** ********************************** ************************
XLL1        ACACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL2        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL3        --ACCAGCACGTTCTCCATCCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL4        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL5        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL6        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL7        -----CAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL8        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL9        --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL10       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL11       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL12       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL13       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL14       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL15       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL16       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL17       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL18       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL19       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL20       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL21       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL22       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL23       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL24       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL25       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLL26       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XLL27       --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XA          --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XC          --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC
XLS         --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAGATGTGATTATTTGCTGGCTAGATAC
XM          --ACCAGCACGTTCTCCATTCCACCGGGAAGGGGTTCCCATGGGAACAGAGCAAAATGTGATTATTTGCTGGCTAGATAC

                ******* ********************************** ********* ****************************
XLL1        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL2        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL3        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL4        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL5        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL6        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL7        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL8        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL9        AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL10       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL11       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL12       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL13       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL14       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL15       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL16       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL17       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL18       AAAGCACGAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL19       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL20       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL21       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL22       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL23       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL24       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL25       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL26       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLL27       AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XA          AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XC          AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCCCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XLS         AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTAGAAATAGCACTAATGTCGGTATAAACAT
XM          AAAGCACAAGCATTTACGCTTTTAATATGCACTAATTATCCTCAAGTAACTGGAAATAGCACTAATGTCGGTATAAACAT
```

```
                    ********************* ** ****** * ********** *******  ****** ************** ********
XLL1        TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL2        TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL3        TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL4        TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL5        TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL6        TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL7        TAATGGAGAAGGATTTTGTAAGAGAAAGGAATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL8        TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL9        TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGAACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT
XLL10       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTTTTCGCAGTCACATTTCGAAGCTT
XLL11       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL12       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL13       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL14       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL15       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL16       TAATGGAGAAGGATTTTGTAAGAAAAAGGGACTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL17       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL18       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL19       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL20       TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL21       TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL22       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL23       TAATGGAGAAGGATTTTGTAAGAGAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL24       TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGAACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT
XLL25       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLL26       TAATGGAGAAGGATTTTGTAAGAAAAAGGGA-TTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTCCGAAGCTT
XLL27       TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XA          TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGGACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT
XC          TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XLS         TAATGGAGAAGGATTTTGTAAGAAAAAGGGATTTCATCCTTGAACAGACATGTATTGTTCGCAGTCACATTTCGAAGCTT
XM          TAATGGAGAAGGATTTTGTATGAAAAAGGGATTTCATCCTTGGACAGAC--GTATTGTTCGCAGTCACATTTCGAAGCTT

                    ********************** * ******************* ******** ******* ** **************
XLL1        GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL2        GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL3        GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL4        GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL5        GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL6        GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTGGTTCTGCCTGCTGCTCCTATAAATT---
XLL7        GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL8        GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL9        GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL10       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL11       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL12       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL13       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL14       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL15       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL16       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL17       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL18       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL19       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL20       GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL21       GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL22       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCTTGCTGCTCCTATAAATT---
XLL23       GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATAGGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL24       GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL25       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLL26       GAAGGTCAAAGTGTCTCACTAGAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATTTCA
XLL27       GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XA          GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XC          GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
XLS         GAAGGTCAAAGTGTCTCACTAAAGGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGTTGCTCCTATAAATTTCA
XM          GAAGGTCAAAGTGTCTCACTAAAAGTCATAATGGACAAAATATGAATTCCTTGTTCTGCCTGCTGCTCCTATAAATT---
```

```
          * ****************************** ************** *** ************ *** ***** *****
XLL1      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL2      CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL3      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL4      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL5      CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL6      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL7      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL8      CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL9      CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL10     CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TACTTTACCTATGTG
XLL11     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL12     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL13     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL14     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL15     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL16     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCCATGTG
XLL17     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL18     CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL19     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL20     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL21     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL22     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTGCCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL23     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL24     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCCCCTAAAACAAAAA-TAC-TTACCTATGTG
XLL25     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL26     CCATTCAAAGGGGGTTTTCTGGCACCCATAGAATATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLL27     CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XA        CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XC        CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XLS       CCATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAA-TAC-TTACCTATGTG
XM        CAATTCAAAGGGGGTTTTCTGGCACCCATAGAGTATCCCTGAAATGTACCCTCTAAAACAAAAAATAC-TTACCTATGTG

           ************************  *********** ****** ** ***************** **************
XLL1      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL2      CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL3      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAGGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL4      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL5      CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL6      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL7      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL8      CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL9      CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL10     CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL11     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL12     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL13     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL14     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL15     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL16     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL17     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL18     CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL19     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL20     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL21     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTTGGGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL22     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL23     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL24     CTGGCATTTTGGACATTTGTAGACACGACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL25     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL26     CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC
XLL27     CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XA        CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
XC        CTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGCCTGTGTATACTGTC
XLS       CTGGCATTTTGGACATTTGTAGACAGAACAGATGATTT-GGGGAAAGATACTGTATGTTAAGGTGACTGTGTATACTGTC
XM        TTGGCATTTTGGACATTTGTAGACACAACAGATGATTT-GGGGAAAGAGACTGTATGTTAAGGTGACTGTGTATACTGTC
```

```
          ** * ** ** ********** ***************** ********** ** ***
XLL1      CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL2      CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGC------------------------
XLL3      CCTTCCCCGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL4      CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL5      CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGC------------------------
XLL6      CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL7      CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL8      CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGC------------------------
XLL9      CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGCCTGGACTGAGAATTAAAATAGGCCC
XLL10     CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGC------------------------
XLL11     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL12     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL13     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL14     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL15     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL16     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL17     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL18     CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGC------------------------
XLL19     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL20     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCGGGC------------------------
XLL21     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL22     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL23     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL24     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------CTGGACTGAGAATTAAAAT
XLL25     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLL26     CCTTCCCTGAGCATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTATCAGGC------------------------
XLL27     CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XA        CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------CTGGACTGAGAAT
XC        CCCTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------------------
XLS       CCTTTCCTGAGCATTAGCAATAGGAGATTTAGTGAAATAGGTCCTGGTATCAGGC------------------------
XM        CCTTCCCTGAACATTAGCAACAGGAGATTTAGTGAAAGAGGTCCTGGTTTCAGGC------------CTGGACTGAGAAT

                                                                  * ******************************
XLL1      ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL2      ----------------------------------------------------CTGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL3      ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL4      ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL5      ----------------------------------------------------CTGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL6      ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL7      ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL8      ----------------------------------------------------CTGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL9      TGGCATTT-------C-----AGGTACACAGAGACCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL10     ----------------------------------------------------CTGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL11     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL12     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL13     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL14     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL15     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL16     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL17     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL18     ----------------------------------------------------CTAGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL19     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL20     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL21     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL22     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL23     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL24     AGGCCCTGGCATT-TC-----AGGTACACAGAGACCCAAACACAGGTCTGGGCTGAGAATTAAAATAGGCCCTGGCATTT
XLL25     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL26     ----------------------------------------------------CTGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLL27     ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XA        TAAAATAGGCCCTGGCATTTCAGGTACACAGAGACCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTT
XC        ----------------------------------------------------CCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XLS       ----------------------------------------------------TCGGACTGAGAATTAAAATAGGCCCTGGCATTT
XM        TAAAATAGGCCCTGGCATTTCAGGTACACAGAGACCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTT
```

```
           *** *********
XLL1       CAGGTACACAGAG-------------------------------------------------------------------------
XLL2       CAGGTACACAGAG-------------------------------------------------------------------------
XLL3       CAGGTACACAGAG-------------------------------------------------------------------------
XLL4       CAGGTACACAGAG-------------------------------------------------------------------------
XLL5       CAGGTACACAGAG-------------------------------------------------------------------------
XLL6       CAGGTACACAGAG-------------------------------------------------------------------------
XLL7       CAGGTACACAGAG-------------------------------------------------------------------------
XLL8       CAGGTACACAGAG-------------------------------------------------------------------------
XLL9       CAGGTACACAGAGGCGCAATCAGC--------------------------------------------------------------
XLL10      CAGGTACACAGAG-------------------------------------------------------------------------
XLL11      CAGGTACACAGAG-------------------------------------------------------------------------
XLL12      CAGGTACACAGAG-------------------------------------------------------------------------
XLL13      CAGGTACACAGAG-------------------------------------------------------------------------
XLL14      CAGGTACACAGAG-------------------------------------------------------------------------
XLL15      CAGGTACACAGAG-------------------------------------------------------------------------
XLL16      CAGGTACACAGAG-------------------------------------------------------------------------
XLL17      CAGGTACACAGAG-------------------------------------------------------------------------
XLL18      CAGGTACACAGAG-------------------------------------------------------------------------
XLL19      CAGGTACACAGAG-------------------------------------------------------------------------
XLL20      CAGGTACACAGAG-------------------------------------------------------------------------
XLL21      CAGGTACACAGAG-------------------------------------------------------------------------
XLL22      CAGGTACACAGAG-------------------------------------------------------------------------
XLL23      CAGGTACACAGAG-------------------------------------------------------------------------
XLL24      CAGGTACACAGAGGCGCAATCAGC--------------------------------------------------------------
XLL25      CAGGTACACAGAG-------------------------------------------------------------------------
XLL26      CAGGTACACAGAG-------------------------------------------------------------------------
XLL27      CAGGTACACAGAG-------------------------------------------------------------------------
XA         CAGGTACACAGAGGCGCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTA----------------------TGG
XC         CAGGTACACAGAG-------------------------------------------------------------------------
XLS        CAGTTACACAGAG-------------------------------------------------------------------------
XM         CAGGTACACAGAGGCGCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGG


XLL1       -------------------------------------------------------------------------------------
XLL2       -------------------------------------------------------------------------------------
XLL3       -------------------------------------------------------------------------------------
XLL4       -------------------------------------------------------------------------------------
XLL5       -------------------------------------------------------------------------------------
XLL6       -------------------------------------------------------------------------------------
XLL7       -------------------------------------------------------------------------------------
XLL8       -------------------------------------------------------------------------------------
XLL9       ----------------------------------------------------------------------CCCCACCAGC----
XLL10      -------------------------------------------------------------------------------------
XLL11      -------------------------------------------------------------------------------------
XLL12      -------------------------------------------------------------------------------------
XLL13      -------------------------------------------------------------------------------------
XLL14      -------------------------------------------------------------------------------------
XLL15      -------------------------------------------------------------------------------------
XLL16      -------------------------------------------------------------------------------------
XLL17      -------------------------------------------------------------------------------------
XLL18      -------------------------------------------------------------------------------------
XLL19      -------------------------------------------------------------------------------------
XLL20      -------------------------------------------------------------------------------------
XLL21      -------------------------------------------------------------------------------------
XLL22      -------------------------------------------------------------------------------------
XLL23      -------------------------------------------------------------------------------------
XLL24      -------------CCCCACCAGC-----CCACTAAATACCGACTTTCTAT-GGCACCTTATAGCAGCCCCTCTGGCATTT
XLL25      -------------------------------------------------------------------------------------
XLL26      -------------------------------------------------------------------------------------
XLL27      -------------------------------------------------------------------------------------
XA         CACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACAGATTGCCAGTCCGGGCCT----------------------
XC         -------------------------------------------------------------------------------------
XLS        -------------------------------------------------------------------------------------
XM         CATTTGCCAGAACTCACAG-------------------------------------------------------------------
```

```
XLL1        ----------------------------------------------------------------------------------------
XLL2        ----------------------------------------------------------------------------------------
XLL3        ----------------------------------------------------------------------------------------
XLL4        ----------------------------------------------------------------------------------------
XLL5        ----------------------------------------------------------------------------------------
XLL6        ----------------------------------------------------------------------------------------
XLL7        ----------------------------------------------------------------------------------------
XLL8        ----------------------------------------------------------------------------------------
XLL9        -CCACTAAATACCGACTTTCTAT-GGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACTCACAGATTGCCAGTCCGG
XLL10       ----------------------------------------------------------------------------------------
XLL11       ----------------------------------------------------------------------------------------
XLL12       ----------------------------------------------------------------------------------------
XLL13       ----------------------------------------------------------------------------------------
XLL14       ----------------------------------------------------------------------------------------
XLL15       ----------------------------------------------------------------------------------------
XLL16       ----------------------------------------------------------------------------------------
XLL17       ----------------------------------------------------------------------------------------
XLL18       ----------------------------------------------------------------------------------------
XLL19       ----------------------------------------------------------------------------------------
XLL20       ----------------------------------------------------------------------------------------
XLL21       ----------------------------------------------------------------------------------------
XLL22       ----------------------------------------------------------------------------------------
XLL23       ----------------------------------------------------------------------------------------
XLL24       GCCAGAACTCACAGATTGCCAGTCCGGGCCT----------------------------------------------------------
XLL25       ----------------------------------------------------------------------------------------
XLL26       ----------------------------------------------------------------------------------------
XLL27       ----------------------------------------------------------------------------------------
XA          ----------------------------------------------------------------------------------------
XC          ----------------------------------------------------------------------------------------
XLS         ----------------------------------------------------------------------------------------
XM          ----------------------------------------------------------------------------------------

                                  ********
XLL1        ----A----------------CCCAAACA----------------------------------------------------------
XLL2        ----G----------------CCCAAACA----------------------------------------------------------
XLL3        ----A----------------CCCAAACA----------------------------------------------------------
XLL4        ----A----------------CCCAAACA----------------------------------------------------------
XLL5        ----G----------------CCCAAACA----------------------------------------------------------
XLL6        ----A----------------CCCAAACA----------------------------------------------------------
XLL7        ----A----------------CCCAAACA----------------------------------------------------------
XLL8        ----G----------------CCCAAACA----------------------------------------------------------
XLL9        GCCTG----------------CCCAAACA----------------------------------------------------------
XLL10       ----G----------------CCCAAACA----------------------------------------------------------
XLL11       ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XLL12       ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XLL13       ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XLL14       ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XLL15       ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XLL16       ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XLL17       ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XLL18       ----G----------------CCCAAACA----------------------------------------------------
XLL19       ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XLL20       ----A----------------CCCAAACA----------------------------------------------------
XLL21       ----A----------------CCCAAACA----------------------------------------------------
XLL22       ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XLL23       ----A----------------CCCAAACA----------------------------------------------------
XLL24       ----G----------------CCCAAACA----------------------------------------------------
XLL25       ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XLL26       ----G----------------CCCAAACA----------------------------------------------------
XLL27       ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XA          ----G----------------CCCAAACA----------------------------------------------------
XC          ----A----------------CCCAAACACAGGTCTGGACTGAGAATTAAAATAGGCCCTGGCATTTCAGGTACACAGA
XLS         ----G----------------CCCAAACA----------------------------------------------------
XM          ----ATTGCCAGTCCGGGCCTGCCCAAACA----------------------------------------------------
```

```
                  ******* ******** **** *** ******************************** ************
XLL1        ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL2        ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGCCATTTGCCAGAA
XLL3        ----------GCCCCACCTGCCCACTAGATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL4        ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL5        ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGCCATTTGCCAGAA
XLL6        ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL7        ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL8        ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGCCATTTGCCAGAA
XLL9        ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL10       ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGCCATTTGCCAGAA
XLL11       GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL12       GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL13       GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL14       GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL15       GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL16       GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL17       GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL18       ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGCCATTTGCCAGAA
XLL19       GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL20       ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL21       ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL22       GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL23       ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL24       ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL25       GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLL26       ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGCCATTTGCCAGAA
XLL27       GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACCTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XA          ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XC          GGCCCAATCAGCCCCCACCAGCCCACTAAATACCGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XLS         ----------GCCCCACCAGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA
XM          ----------GCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAA

                  * ***
XLL1        CCCAC-------------------------------------------------------------------------
XLL2        CCCAC-------------------------------------------------------------------------
XLL3        CCCAC-------------------------------------------------------------------------
XLL4        CCCAC-------------------------------------------------------------------------
XLL5        CCCAC-------------------------------------------------------------------------
XLL6        CCCAC-------------------------------------------------------------------------
XLL7        CCCAC-------------------------------------------------------------------------
XLL8        CCCAC-------------------------------------------------------------------------
XLL9        CCCAC-------------------------------------------------------------------------
XLL10       CCCAC-------------------------------------------------------------------------
XLL11       CTCACAGATTGCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCA
XLL12       CTCAC-------------------------------------------------------------------AGATT
XLL13       CTCAC-----------------------------------------------------------------AGATT
XLL14       CTCAC---------------------------------------------------------------AGATT
XLL15       CTCAC-------------------------------------------------------------------AGATT
XLL16       CTCAC-----------------------------------------------------------------AGATT
XLL17       CTCAC-----------------------------------------------------------------AGATT
XLL18       CCCAC-------------------------------------------------------------------------
XLL19       CTCAC-----------------------------------------------------------------AGATT
XLL20       CCCAC-------------------------------------------------------------------------
XLL21       CCCAC-------------------------------------------------------------------------
XLL22       CTCAC-----------------------------------------------------------------AGATT
XLL23       CCCAC-------------------------------------------------------------------------
XLL24       CCCAC-------------------------------------------------------------------------
XLL25       CTCAC-----------------------------------------------------------------AGATT
XLL26       CCCAC-------------------------------------------------------------------------
XLL27       CTCAC-------------------------------------------------------------------AGATT
XA          CCCAC-------------------------------------------------------------------------
XC          CTCAC-------------------------------------------------------------------------
XLS         CCCAC-------------------------------------------------------------------------
XM          CCCAC-------------------------------------------------------------------------
```

```
XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    ----------------------------------------------------------------------------------------
XLL5    ----------------------------------------------------------------------------------------
XLL6    ----------------------------------------------------------------------------------------
XLL7    ----------------------------------------------------------------------------------------
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   GCCCCTCTGGCATTTGCCAGA----------ACCCAC----------------------------------------------------
XLL12   GCCAGTCCGGGCCTGCCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGG
XLL13   GCCAGTCCGGGCCTGCCCAAACAGC----------------------------------------------------------------
XLL14   GCCAGTCCGGGCCTGCCCAAACAGC----------------------------------------------------------------
XLL15   GCCAGTCCGGGCCTGCCCAAACAGC----------------------------------------------------------------
XLL16   GCCAGTCCGGGCCTGCCCAAACAGC----------------------------------------------------------------
XLL17   GCCAGTCCGGGCCTGCCCAAACAGC----------------------------------------------------------------
XLL18   ----------------------------------------------------------------------------------------
XLL19   GCCAGTCCGGGCCTGCCCAAACAGC----------------------------------------------------------------
XLL20   ----------------------------------------------------------------------------------------
XLL21   ----------------------------------------------------------------------------------------
XLL22   GCCAGTCCGGGCCTGCCCAAACAGC----------------------------------------------------------------
XLL23   ----------------------------------------------------------------------------------------
XLL24   ----------------------------------------------------------------------------------------
XLL25   GCCAGTCCGGGCCTGCCCAAACAGC----------------------------------------------------------------
XLL26   ----------------------------------------------------------------------------------------
XLL27   GCCAGTCCGGGCCTGCCCAAACAGC----------------------------------------------------------------
XA      ----------------------------------------------------------------------------------------
XC      ----------------------------------------------------------------------------------------
XLS     ----------------------------------------------------------------------------------------
XM      ----------------------------------------------------------------------------------------


XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    ----------------------------------------------------------------------------------------
XLL5    ----------------------------------------------------------------------------------------
XLL6    ----------------------------------------------------------------------------------------
XLL7    ----------------------------------------------------------------------------------------
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   ----------------------------------------------------------------------------------------
XLL12   CATTTGCCAGAACCCAC----------------------------------------------------------------------
XLL13   ----------CCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAAC
XLL14   ----------------------------------------------------------------------------------C
XLL15   ----------------------------------------------------------------------------------------
XLL16   ----------------------------------------------------------------------------------------
XLL17   ----------------------------------------------------------------------------------------
XLL18   ----------------------------------------------------------------------------------------
XLL19   ----------------------------------------------------------------------------------------
XLL20   ----------------------------------------------------------------------------------------
XLL21   ----------------------------------------------------------------------------------------
XLL22   ----------------------------------------------------------------------------------------
XLL23   ----------------------------------------------------------------------------------------
XLL24   ----------------------------------------------------------------------------------------
XLL25   ----------------------------------------------------------------------------------------
XLL26   ----------------------------------------------------------------------------------------
XLL27   ----------------------------------------------------------------------------------------
XA      ----------------------------------------------------------------------------------------
XC      ----------------------------------------------------------------------------------------
XLS     ----------------------------------------------------------------------------------------
XM      ----------------------------------------------------------------------------------------
```

```
XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    ----------------------------------------------------------------------------------------
XLL5    ----------------------------------------------------------------------------------------
XLL6    ----------------------------------------------------------------------------------------
XLL7    ----------------------------------------------------------------------------------------
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   ----------------------------------------------------------------------------------------
XLL12   ----------------------------------------------------------------------------------------
XLL13   CCAC------------------------------------------------------------------------------------
XLL14   CCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCAC---------
XLL15   ----------------------------------------------------------CCCACCTGCCCACT
XLL16   ----------------------------------------------------------------------------------------
XLL17   ----------------------------------------------------------------------------------------
XLL18   ----------------------------------------------------------------------------------------
XLL19   ----------------------------------------------------------------------------------------
XLL20   ----------------------------------------------------------------------------------------
XLL21   ----------------------------------------------------------------------------------------
XLL22   ----------------------------------------------------------------------------------------
XLL23   ----------------------------------------------------------------------------------------
XLL24   ----------------------------------------------------------------------------------------
XLL25   ----------------------------------------------------------------------------------------
XLL26   ----------------------------------------------------------------------------------------
XLL27   ----------------------------------------------------------------------------------------
XA      ----------------------------------------------------------------------------------------
XC      ----------------------------------------------------------------------------------------
XLS     ----------------------------------------------------------------------------------------
XM      ----------------------------------------------------------------------------------------


XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    ----------------------------------------------------------------------------------------
XLL5    ----------------------------------------------------------------------------------------
XLL6    ----------------------------------------------------------------------------------------
XLL7    ----------------------------------------------------------------------------------------
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   ----------------------------------------------------------------------------------------
XLL12   ----------------------------------------------------------------------------------------
XLL13   ----------------------------------------------------------------------------------------
XLL14   ----------------------------------------------------------------------------------------
XLL15   AAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCAC----------------------
XLL16   ----------------------------------------------------CCCACCTGCCCACTAAATACTGACTTT
XLL17   ----------------------------------------------------------------------------------------
XLL18   ----------------------------------------------------------------------------------------
XLL19   ----------------------------------------------------------------------------------------
XLL20   ----------------------------------------------------------------------------------------
XLL21   ----------------------------------------------------------------------------------------
XLL22   ----------------------------------------------------------------------------------------
XLL23   ----------------------------------------------------------------------------------------
XLL24   ----------------------------------------------------------------------------------------
XLL25   ----------------------------------------------------------------------------------------
XLL26   ----------------------------------------------------------------------------------------
XLL27   ----------------------------------------------------------------------------------------
XA      ----------------------------------------------------------------------------------------
XC      ----------------------------------------------------------------------------------------
XLS     ----------------------------------------------------------------------------------------
XM      ----------------------------------------------------------------------------------------
```

```
XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    ----------------------------------------------------------------------------------------
XLL5    ----------------------------------------------------------------------------------------
XLL6    ----------------------------------------------------------------------------------------
XLL7    ----------------------------------------------------------------------------------------
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   ----------------------------------------------------------------------------------------
XLL12   ----------------------------------------------------------------------------------------
XLL13   ----------------------------------------------------------------------------------------
XLL14   ----------------------------------------------------------------------------------------
XLL15   ----------------------------------------------------------------------------------------
XLL16   CTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCAC-----------------------------------------
XLL17   ----------------------------------------CCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTA
XLL18   ----------------------------------------------------------------------------------------
XLL19   ----------------------------------------------------------------------------------------
XLL20   ----------------------------------------------------------------------------------------
XLL21   ----------------------------------------------------------------------------------------
XLL22   ----------------------------------------------------------------------------------------
XLL23   ----------------------------------------------------------------------------------------
XLL24   ----------------------------------------------------------------------------------------
XLL25   ----------------------------------------------------------------------------------------
XLL26   ----------------------------------------------------------------------------------------
XLL27   ----------------------------------------------------------------------------------------
XA      ----------------------------------------------------------------------------------------
XC      ----------------------------------------------------------------------------------------
XLS     ----------------------------------------------------------------------------------------
XM      ----------------------------------------------------------------------------------------


XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    ----------------------------------------------------------------------------------------
XLL5    ----------------------------------------------------------------------------------------
XLL6    ----------------------------------------------------------------------------------------
XLL7    ----------------------------------------------------------------------------------------
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   ----------------------------------------------------------------------------------------
XLL12   ----------------------------------------------------------------------------------------
XLL13   ----------------------------------------------------------------------------------------
XLL14   ----------------------------------------------------------------------------------------
XLL15   ----------------------------------------------------------------------------------------
XLL16   ----------------------------------------------------------------------------------------
XLL17   TAGCAGCCCCTCTGGCATTTGCCAGAACCCAC------------------------------------------------------
XLL18   ----------------------------------------------------------------------------------------
XLL19   ----------------------------CCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCT
XLL20   ----------------------------------------------------------------------------------------
XLL21   ----------------------------------------------------------------------------------------
XLL22   ----------------------------------------------------------------------------------------
XLL23   ----------------------------------------------------------------------------------------
XLL24   ----------------------------------------------------------------------------------------
XLL25   ----------------------------------------------------------------------------------------
XLL26   ----------------------------------------------------------------------------------------
XLL27   ----------------------------------------------------------------------------------------
XA      ----------------------------------------------------------------------------------------
XC      ----------------------------------------------------------------------------------------
XLS     ----------------------------------------------------------------------------------------
XM      ----------------------------------------------------------------------------------------
```

```
                                                                                ************   * ***
XLL1      ------------------------------------------------------------AGATTGCCAGTCTGGGCCTG
XLL2      ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL3      ------------------------------------------------------------AGATTGCCAGTCTGGGCCTG
XLL4      ------------------------------------------------------------AGATTGCCAGTCTGGGTCTG
XLL5      ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL6      ------------------------------------------------------------AGATTGCCAGTCTGGGCCTG
XLL7      ------------------------------------------------------------AGATTGCCAGTCTGGGCCTG
XLL8      ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL9      ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL10     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL11     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL12     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL13     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL14     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL15     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL16     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL17     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL18     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL19     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL20     ------------------------------------------------------------AGATTGCCAGTCTGGGCCTG
XLL21     ------------------------------------------------------------AGATTGCCAGTCTGGGCCTG
XLL22     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL23     ------------------------------------------------------------AGATTGCCAGTCTGGGCCTG
XLL24     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL25     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL26     ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLL27     CTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAATCCACAGATTGCCAGTCCGGGCCTG
XA        ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XC        ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG
XLS       ------------------------------------------------------------AGATTGCCAGTCCAAGCCTG
XM        ------------------------------------------------------------AGATTGCCAGTCCGGGCCTG

          *                                                            **** ***
XLL1      C------------------------------------------------------------CTGGTATT-----------
XLL2      C------------------------------------------------------------CTGGTATT-----------
XLL3      C------------------------------------------------------------CTGGTATT-----------
XLL4      C------------------------------------------------------------CTGGTATT-----------
XLL5      C------------------------------------------------------------CTGGTATT-----------
XLL6      C------------------------------------------------------------CTGGTATT-----------
XLL7      C------------------------------------------------------------CTGGTATT-----------
XLL8      C------------------------------------------------------------CTGGTATT-----------
XLL9      C------------------------------------------------------------CTGGTATT-----------
XLL10     C------------------------------------------------------------CTGGTATT-----------
XLL11     C------------------------------------------------------------CTGGTATT-----------
XLL12     C------------------------------------------------------------CTGGTATT-----------
XLL13     C------------------------------------------------------------CTGGTATT-----------
XLL14     C------------------------------------------------------------CTGGTATT-----------
XLL15     C------------------------------------------------------------CTGGTATT-----------
XLL16     C------------------------------------------------------------CTGGTATT-----------
XLL17     C------------------------------------------------------------CTGGTATT-----------
XLL18     C------------------------------------------------------------CTGGTATT-----------
XLL19     C------------------------------------------------------------CTGGTATT-----------
XLL20     C------------------------------------------------------------CTGGTATT-----------
XLL21     C------------------------------------------------------------CTGGTATT-----------
XLL22     C------------------------------------------------------------CTGGTATT-----------
XLL23     C------------------------------------------------------------CTGGTATT-----------
XLL24     C------------------------------------------------------------CTGGTATT-----------
XLL25     C------------------------------------------------------------CTGGTATT-----------
XLL26     C------------------------------------------------------------CTGGTATT-----------
XLL27     C------------------------------------------------------------CTGGTATT-----------
XA        C------------------------------------------------------------CTGGTATT-----------
XC        CCCAAACAGCCCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACCC
XLS       C------------------------------------------------------------CTGGTATT-----------
XM        C------------------------------------------------------------CTGGTATT-----------
```

```
XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    ----------------------------------------------------------------------------------------
XLL5    ----------------------------------------------------------------------------------------
XLL6    ----------------------------------------------------------------------------------------
XLL7    ----------------------------------------------------------------------------------------
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   ----------------------------------------------------------------------------------------
XLL12   ----------------------------------------------------------------------------------------
XLL13   ----------------------------------------------------------------------------------------
XLL14   ----------------------------------------------------------------------------------------
XLL15   ----------------------------------------------------------------------------------------
XLL16   ----------------------------------------------------------------------------------------
XLL17   ----------------------------------------------------------------------------------------
XLL18   ----------------------------------------------------------------------------------------
XLL19   GGCATTTGCCAGAACCCAC-------------------------------------------------------------------
XLL20   ----------------------------------------------------------------------------------------
XLL21   ----------------------------------------------------------------------------------------
XLL22   --------------CCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGA
XLL23   ----------------------------------------------------------------------------------------
XLL24   ----------------------------------------------------------------------------------------
XLL25   ----------------------------------------------------------------------------------------
XLL26   ----------------------------------------------------------------------------------------
XLL27   ----------------------------------------------------------------------------------------
XA      ----------------------------------------------------------------------------------------
XC      ----------------------------------------------------------------------------------------
XLS     ----------------------------------------------------------------------------------------
XM      ----------------------------------------------------------------------------------------


XLL1    ----------------------------------------------------------------------------------------
XLL2    ----------------------------------------------------------------------------------------
XLL3    ----------------------------------------------------------------------------------------
XLL4    ----------------------------------------------------------------------------------------
XLL5    ----------------------------------------------------------------------------------------
XLL6    ----------------------------------------------------------------------------------------
XLL7    ----------------------------------------------------------------------------------------
XLL8    ----------------------------------------------------------------------------------------
XLL9    ----------------------------------------------------------------------------------------
XLL10   ----------------------------------------------------------------------------------------
XLL11   ----------------------------------------------------------------------------------------
XLL12   ----------------------------------------------------------------------------------------
XLL13   ----------------------------------------------------------------------------------------
XLL14   ----------------------------------------------------------------------------------------
XLL15   ----------------------------------------------------------------------------------------
XLL16   ----------------------------------------------------------------------------------------
XLL17   ----------------------------------------------------------------------------------------
XLL18   ----------------------------------------------------------------------------------------
XLL19   ----------------------------------------------------------------------------------------
XLL20   ----------------------------------------------------------------------------------------
XLL21   ----------------------------------------------------------------------------------------
XLL22   ACCCAC--------------------------------------------------------------------------------
XLL23   ----------------------------------------------------------------------------------------
XLL24   ----------------------------------------------------------------------------------------
XLL25   -CCCACCTGCCCACTAAATACTGACTTTCTATGGCACCTTATAGCAGCCCCTCTGGCATTTGCCAGAACCCAC-------
XLL26   ----------------------------------------------------------------------------------------
XLL27   ------------------------------------------------------------------------CCCACCTGCCCA
XA      ----------------------------------------------------------------------------------------
XC      ----------------------------------------------------------------------------------------
XLS     ----------------------------------------------------------------------------------------
XM      ----------------------------------------------------------------------------------------
```

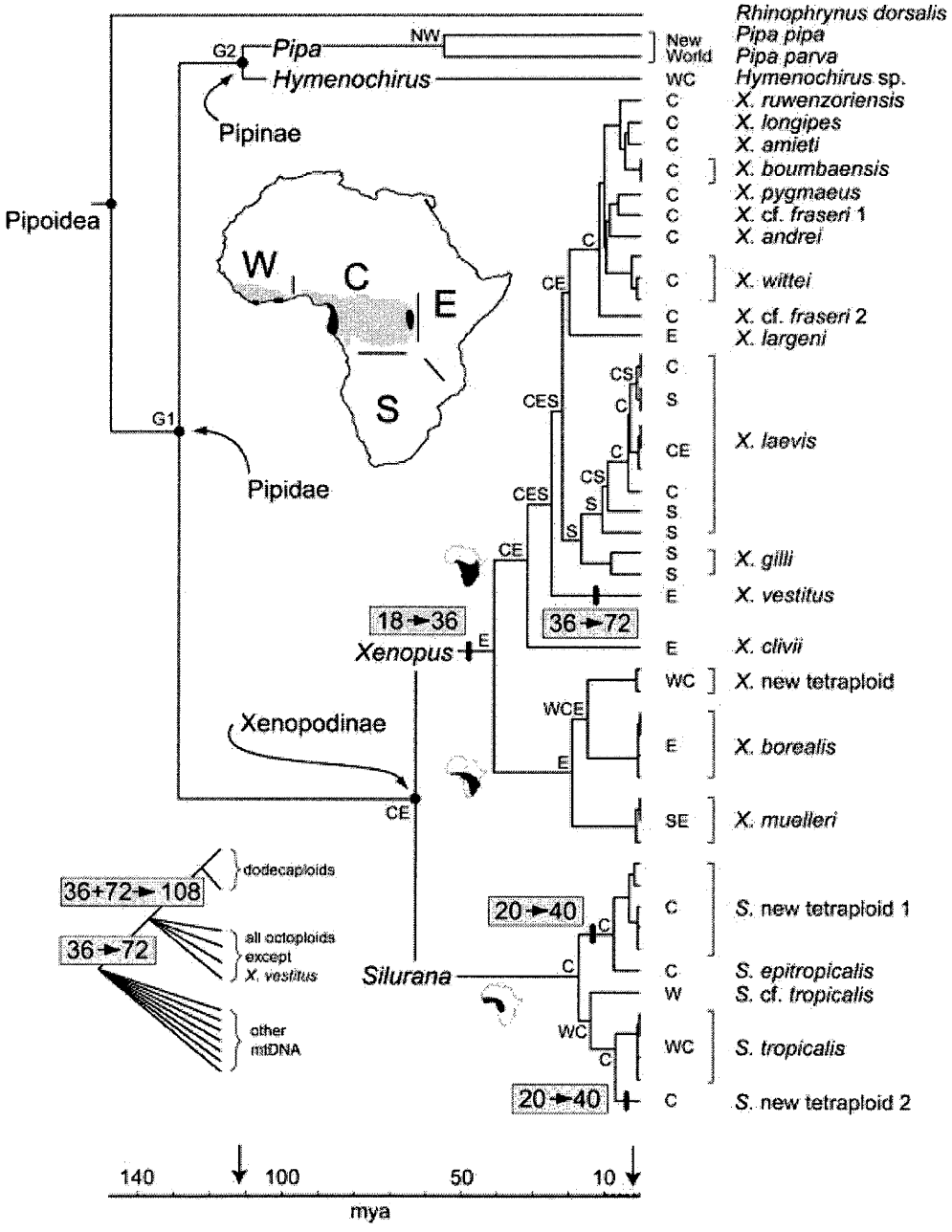
```
                                        ** ********************* ********** ***************
XLL1    --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL2    --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL3    --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTTAACAGACTGTGCCT
XLL4    --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL5    --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL6    --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL7    --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL8    --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL9    --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL10   --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL11   --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL12   --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL13   --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL14   --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL15   --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL16   --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL17   --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL18   --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL19   --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL20   --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL21   --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL22   --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL23   --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL24   --------------------------------AACCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL25   --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLL26   --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XLL27   --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XA      --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XC      ACAGATTGCCAGTCCGGGCCTGCCTGGTATTAATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT
XLS     --------------------------------AATCATACATCCCTTCTCTATCTATTTGCTCAGTCAACAGACTGTGCCT
XM      --------------------------------AATCATACATCCCTTCTCTATCTGTTTGCTCAGTCAACAGACTGTGCCT

        *********** ****************************************** ***************************
XLL1    GTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL2    GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL3    GTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL4    GTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL5    GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL6    GTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL7    GTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL8    GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL9    GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL10   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL11   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL12   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL13   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL14   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL15   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL16   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL17   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL18   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL19   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL20   GTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL21   GTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL22   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL23   GTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL24   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL25   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL26   GTAGGCAGTTACAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLL27   GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XA      GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XC      GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
XLS     GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCATTTGATTAGTGGAGTGGAATCTGTGTG
XM      GTAGGCAGTTAAAATCCATTGACACATTCCTAATGGCTCGTTCATTCGCAGCAGTTGATTAGTGGAGTGGAATCTGTGTG
```

```
          ***** ************************************ ******* ********** ***************** *
XLL1      ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL2      ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL3      ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL4      ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL5      ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL6      ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL7      ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL8      ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL9      ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL10     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL11     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL12     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL13     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL14     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL15     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL16     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCCG
XLL17     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL18     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL19     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL20     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL21     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCACGGCTGCTTTCCATTACCAG
XLL22     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL23     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACGTTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL24     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL25     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL26     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLL27     ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XA        ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XC        ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG
XLS       ATCAAGTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTCTTAACTTGCATGGCTGCTTTCCATTACCAG
XM        ATCAACTCAATAAGTGGTTACCTCTGCATCAATTGATGGACATTAGGTTTTTAACTTGCATGGCTGCTTTCCATTACCAG

          ************************************ * ************ ** ***************************
XLL1      AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL2      AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCCTGTCATATAAATAGTCTTTCCACTGGCCT
XLL3      AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL4      AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL5      AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL6      AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCTCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL7      AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL8      AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCCTGTCATATAAATAGTCTTTCCACTGGCCT
XLL9      AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL10     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCCTGTCATATAAATAGTCTTTCCACTGGCCT
XLL11     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL12     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL13     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL14     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL15     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL16     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACCCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL17     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL18     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCCTGTCATATAAATAGTCTTTCCACTGGCCT
XLL19     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL20     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL21     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL22     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL23     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL24     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL25     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLL26     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCCTGTCATATAAATAGTCTTTCCACTGGCCT
XLL27     AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XA        AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XC        AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
XLS       AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGCCATATAAATAGTCTTTCCACTGGCCT
XM        AGTGGGTGTCTCCTGATGTCACTCACAACACAAACTCCCAAAGGCAGGCCATGTCATATAAATAGTCTTTCCACTGGCCT
```

```
          ** **********  ****** ******** ************ ****************************************
XLL1      CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL2      CATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL3      CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL4      CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL5      CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL6      CATTAGAAGGTTGTATTAGTGGGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL7      CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL8      CATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL9      CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL10     CATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL11     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL12     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL13     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL14     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL15     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL16     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL17     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL18     CATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCCATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL19     CACTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL20     CATTAGAAGGTTGCATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL21     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL22     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL23     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL24     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL25     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL26     CATTAGAAGGTTGTTTTAGTGAGCTAAGATTTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLL27     CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XA        CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XC        CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XLS       CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
XM        CATTAGAAGGTTGTATTAGTGAGCTAAGATCTCCATCTGATCTATTGCTTGATAGCCTGACCATGCTCAGGAAAGAGAAC
                          +1

          *******
XLL1      ACCCTGCCA
XLL2      ACCCTGCC-
XLL3      ACCCTGCC-
XLL4      ACCCTGCC-
XLL5      ACCCTGC--
XLL6      ACCCTGCC-
XLL7      ACCCTGCC-
XLL8      ACCCTGCC-
XLL9      ACCCTGCC-
XLL10     ACCCTGCC-
XLL11     ACCCTGCC-
XLL12     ACCCTGCC-
XLL13     ACCCTGC--
XLL14     ACCCTGC--
XLL15     ACCCTGC--
XLL16     ACCCTGC--
XLL17     ACCCTGC--
XLL18     ACCCTGC--
XLL19     ACCCTGC--
XLL20     ACCCTGCC-
XLL21     ACCCTGC--
XLL22     ACCCTGCC-
XLL23     ACCCTGCC-
XLL24     ACCCTGCC-
XLL25     ACCCTGCC-
XLL26     ACCCTGCC-
XLL27     ACCCTGCC-
XA        ACCCTGCC-
XC        ACCCTGCC-
XLS       ACCCTGCC-
XM        ACCCTGCC-
```

Xenopus Phylogeny

The phylogenetic relationship of African clawed frogs was investigated by Evans et al (2004) and reproduced here. A phylogenetic tree is provided where branch lengths are proportional to divergence time as estimated from the maximum likelihood topology and a relaxed molecular clock (Evans et al., 2004).



WORKS CITED

Chimpanzee Sequencing and Analysis Consortium. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69-87.

Allain, A. E., Bairi, A., Meyrand, P. and Branchereau, P. (2006). Expression of the glycinergic system during the course of embryonic development in the mouse spinal cord and its co-localization with GABA immunoreactivity. *J Comp Neurol* **496**, 832-46.

Allendorf, F., Thorgaard, G. (1984). Tetraploidy and the evolution of salmonid fishes. In *Evolutionary genetics of fishes*, (ed. B. J. Turner). New York: Plenum Press.

Allendorf, F. W., Knudsen, K. L. and Leary, R. F. (1983). Adaptive significance of differences in the tissue-specific expression of a phosphoglucomutase gene in rainbow trout. *Proc Natl Acad Sci U S A* **80**, 1397-400.

Allendorf, F. W., Knudsen, K. L. and Phelps, S. R. (1982). Identification of a gene regulating the tissue expression of a phosphoglucomutase locus in rainbow trout. *Genetics* **102**, 259-68.

Arata, S., Hashizume, C., Kikusui, T., Takeuchi, Y. and Mori, Y. (2008). Sequences of canine glutamate decarboxylase (GAD) 1 and GAD2 genes, and variation of their genetic polymorphisms among five dog breeds. *J Vet Med Sci* **70**, 1107-10.

Balhoff, J. P. and Wray, G. A. (2005). Evolutionary analysis of the well characterized endo16 promoter reveals substantial variation within functional sites. *Proc Natl Acad Sci U S A* **102**, 8591-6.

Behar, T., Schaffner, A., Laing, P., Hudson, L., Komoly, S. and Barker, J. (1993). Many spinal cord cells transiently express low molecular weight forms of glutamic acid decarboxylase during embryonic development. *Brain Res Dev Brain Res* **72**, 203-18.

Beldade, P., Brakefield, P. M. and Long, A. D. (2002). Contribution of Distal-less to quantitative variation in butterfly eyespots. *Nature* **415**, 315-8.

Blanchette, M., Kent, W. J., Riemer, C., Elnitski, L., Smit, A. F., Roskin, K. M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E. D. et al. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* **14**, 708-15.

Bouwmeester, T. (2001). The Spemann-Mangold organizer: the control of fate specification and morphogenetic rearrangements during gastrulation in Xenopus. *Int J Dev Biol* **45**, 251-8.

Brem, R. B., Yvert, G., Clinton, R. and Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752-5.

Britten, R. J. and Davidson, E. H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev Biol* **46**, 111-38.

Brox, A., Puelles, L., Ferreiro, B. and Medina, L. (2003). Expression of the genes GAD67 and Distal-less-4 in the forebrain of Xenopus laevis confirms a common pattern in tetrapods. *J Comp Neurol* **461**, 370-93.

Bu, D. F., Erlander, M. G., Hitz, B. C., Tillakaratne, N. J., Kaufman, D. L., Wagner-McPherson, C. B., Evans, G. A. and Tobin, A. J. (1992). Two human glutamate decarboxylases, 65-kDa GAD and 67-kDa GAD, are each encoded by a single gene. *Proc Natl Acad Sci U S A* **89**, 2115-9.

Buddhala, C., Hsu, C. C. and Wu, J. Y. (2009). A novel mechanism for GABA synthesis and packaging into synaptic vesicles. *Neurochem Int* **55**, 9-12.
Bush, R. M. and Paigen, K. (1992). Evolution of beta-glucuronidase regulation in the genus Mus. *Evolution* **46**, 1-15.
Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M. T., Glanowski, S., Tanenbaum, D. M., White, T. J., Sninsky, J. J., Hernandez, R. D. et al. (2005). Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153-7.
Cannatella, D. C. and Sa, R. O. D. (1993). Xenopus laevis as a Model Organism. *Systematic Biology* **42**, 476-507.
Cannatella, D. C. and Trueb, L. (1988). Evolution of Pipoid Frogs: Morphology and Phylogenetic Relationships of Pseudhymenochirus. *Journal of Herpetology* **22**, 439-456.
Carey, M. F., Peterson, C. L. and Smale, S. T. (2009). Transcriptional Regulation in Eukaryotes. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
Carlini, D. B. and Stephan, W. (2003). In vivo introduction of unpreferred synonymous codons into the Drosophila Adh gene results in reduced levels of ADH protein. *Genetics* **163**, 239-43.
Carroll, S. B. (1995). Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479-85.
Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25-36.
Carroll, S. B., Grenier, J.K., Weatherbee, S.D. (2001). From DNA to diversity: molecular genetics and the evolution of animal design. Malden, Mass: Blackwell Science.
Celniker, S. E., Sharma, S., Keelan, D. J. and Lewis, E. B. (1990). The molecular genetics of the bithorax complex of Drosophila: cis-regulation in the Abdominal-B domain. *Embo J* **9**, 4277-86.
Chain, F. J. and Evans, B. J. (2006). Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog Xenopus laevis. *PLoS Genet* **2**, e56.
Clark, A. G., Glanowski, S., Nielsen, R., Thomas, P. D., Kejariwal, A., Todd, M. A., Tanenbaum, D. M., Civello, D., Lu, F., Murphy, B. et al. (2003). Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**, 1960-3.
Costa, P. a. P., C. (1999). Genetic analysis of needle proteins in maritime pine. *Silvae Genetics* **48**, 146-150.
Crawford, D. L., Segal, J. A. and Barnett, J. L. (1999). Evolutionary analysis of TATA-less proximal promoter function. *Mol Biol Evol* **16**, 194-207.
Daborn, P. J., Yen, J. L., Bogwitz, M. R., Le Goff, G., Feil, E., Jeffers, S., Tijet, N., Perry, T., Heckel, D., Batterham, P. et al. (2002). A single p450 allele associated with insecticide resistance in Drosophila. *Science* **297**, 2253-6.
Damerval, C., Maurice, A., Josse, J. M. and de Vienne, D. (1994). Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* **137**, 289-301.
de Sa, R. O. and Hillis, D. M. (1990). Phylogenetic relationships of the pipid frogs Xenopus and Silurana: an integration of ribosomal DNA and morphology. *Mol Biol Evol* **7**, 365-76.

De Vienne, D., Bost, B., Fievet, J., Zivy, M., and Dillmann, C. (2001). genetic variability of proteome expression and metabolic control. *Plant Physiology and Biochemistry* **39**, 271-283.
Dehal, P. and Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* **3**, e314.
Depaulis, F., Mousset, S. and Veuille, M. (2001). Haplotype tests using coalescent simulations conditional on the number of segregating sites. *Mol Biol Evol* **18**, 1136-8.
Desmond, M. E. (1982). Description of the occlusion of the spinal cord lumen in early human embryos. *Anat Rec* **204**, 89-93.
Desmond, M. E. and Levitan, M. L. (2002). Brain expansion in the chick embryo initiated by experimentally produced occlusion of the spinal neurocoel. *Anat Rec* **268**, 147-59.
Doniger, S. W. and Fay, J. C. (2007). Frequent gain and loss of functional transcription factor binding sites. *PLoS Comput Biol* **3**, e99.
Dorus, S., Vallender, E. J., Evans, P. D., Anderson, J. R., Gilbert, S. L., Mahowald, M., Wyckoff, G. J., Malcom, C. M. and Lahn, B. T. (2004). Accelerated evolution of nervous system genes in the origin of Homo sapiens. *Cell* **119**, 1027-40.
Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-7.
Ericson, J., Morton, S., Kawakami, A., Roelink, H. and Jessell, T. M. (1996). Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. *Cell* **87**, 661-73.
Erlander, M. G., Tillakaratne, N. J., Feldblum, S., Patel, N. and Tobin, A. J. (1991). Two genes encode distinct glutamate decarboxylases. *Neuron* **7**, 91-100.
Evans, B. J. (2008). Genome evolution and speciation genetics of clawed frogs (Xenopus and Silurana). *Front Biosci* **13**, 4687-706.
Evans, B. J., Kelley, D. B., Melnick, D. J. and Cannatella, D. C. (2005). Evolution of RAG-1 in polyploid clawed frogs. *Mol Biol Evol* **22**, 1193-207.
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**, 197-213.
Fay, J. C. and Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405-13.
Fisher, R. A. (1930). The Genetical Theory of Natural Selection: Clarendon Press.
Fitch, D. H. (1997). Evolution of male tail development in rhabditid nematodes related to Caenorhabditis elegans. *Syst Biol* **46**, 145-79.
Fu, Y. X. (1995). Statistical properties of segregating sites. *Theor Popul Biol* **48**, 172-97.
Fu, Y. X. and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* **133**, 693-709.
Gazave, E., Marques-Bonet, T., Fernando, O., Charlesworth, B. and Navarro, A. (2007). Patterns and rates of intron divergence between humans and chimpanzees. *Genome Biol* **8**, R21.
Gerber, S., Fabre, F., Planchon, C. (2000). Genetics of seed quality in soybean analysed by capillary gel electrophoresis. *Plant Science* **152**, 181-189.

Glazko, G., Veeramachaneni, V., Nei, M. and Makalowski, W. (2005). Eighty percent of proteins are different between humans and chimpanzees. *Gene* **346**, 215-9.
Grens, A., Mason, E., Marsh, J. L. and Bode, H. R. (1995). Evolutionary conservation of a cell fate specification gene: the Hydra achaete-scute homolog has proneural activity in Drosophila. *Development* **121**, 4027-35.
Guirado, S., Davila, J. C., Real, M. A. and Medina, L. (1999). Nucleus accumbens in the lizard Psammodromus algirus: chemoarchitecture and cortical afferent connections. *J Comp Neurol* **405**, 15-31.
Haddrill, P. R., Bachtrog, D. and Andolfatto, P. (2008). Positive and negative selection on noncoding DNA in Drosophila simulans. *Mol Biol Evol* **25**, 1825-34.
Haddrill, P. R., Charlesworth, B., Halligan, D. L. and Andolfatto, P. (2005). Patterns of intron sequence evolution in Drosophila are dependent upon length and GC content. *Genome Biol* **6**, R67.
Halder, G., Callaerts, P. and Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science* **267**, 1788-92.
Halligan, D. L., Eyre-Walker, A., Andolfatto, P. and Keightley, P. D. (2004). Patterns of evolutionary constraints in intronic and intergenic DNA of Drosophila. *Genome Res* **14**, 273-9.
Haygood, R., Fedrigo, O., Hanson, B., Yokoyama, K. D. and Wray, G. A. (2007). Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat Genet* **39**, 1140-4.
Hein, J., Schierup, M. H. and Wiuf, C. (2005). Gene Genealogies, Variation and Evolution - A primer in coalescent theory: Oxford University Press.
Hellmann, I., Zollner, S., Enard, W., Ebersberger, I., Nickel, B. and Paabo, S. (2003). Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res* **13**, 831-7.
Hill, W. G. (1975). Linkage disequilibrium among multiple neutral alleles produced by mutation in finite population. *Theor Popul Biol* **8**, 117-26.
Hoekstra, H. E. and Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution Int J Org Evolution* **61**, 995-1016.
Hudson, R. R. (2001). Two-locus sampling distributions and their application. *Genetics* **159**, 1805-17.
Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337-8.
Hudson, R. R., Kreitman, M. and Aguade, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153-9.
Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**, 318-56.
Johnson, M. B., Kawasawa, Y. I., Mason, C. E., Krsnik, Z., Coppola, G., Bogdanovic, D., Geschwind, D. H., Mane, S. M., State, M. W. and Sestan, N. (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494-509.
Katarova, Z., Sekerkova, G., Prodan, S., Mugnaini, E. and Szabo, G. (2000). Domain-restricted expression of two glutamic acid decarboxylase genes in midgestation mouse embryos. *J Comp Neurol* **424**, 607-27.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-66.
Kaufman, D. L., Houser, C. R. and Tobin, A. J. (1991). Two forms of the gamma-aminobutyric acid synthetic enzyme glutamate decarboxylase have distinct intraneuronal distributions and cofactor interactions. *J Neurochem* **56**, 720-3.
Keightley, P. D., Lercher, M. J. and Eyre-Walker, A. (2005). Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol* **3**, e42.
Kim, J., He, X. and Sinha, S. (2009). Evolution of regulatory sequences in 12 Drosophila species. *PLoS Genet* **5**, e1000330.
King, M. C. and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science* **188**, 107-16.
Kliebenstein, D. J. (2009). Quantification of variation in expression networks. *Methods Mol Biol* **553**, 227-45.
Kobel, H. R. (1996). The Biology of Xenopus: Clardenon Press.
Kuhner, M. K., Yamato, J. and Felsenstein, J. (2000). Maximum likelihood estimation of recombination rates from population data. *Genetics* **156**, 1393-401.
Kusserow, A., Pang, K., Sturm, C., Hrouda, M., Lentfer, J., Schmidt, H. A., Technau, U., von Haeseler, A., Hobmayer, B., Martindale, M. Q. et al. (2005). Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* **433**, 156-60.
Lamigeon, C., Bellier, J. P., Sacchettoni, S., Rujano, M. and Jacquemont, B. (2001). Enhanced neuronal protection from oxidative stress by coculture with glutamic acid decarboxylase-expressing astrocytes. *J Neurochem* **77**, 598-606.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R. et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-8.
Lawniczak, M. K., Holloway, A. K., Begun, D. J. and Jones, C. D. (2008). Genomic analysis of the relationship between gene expression variation and DNA polymorphism in Drosophila simulans. *Genome Biol* **9**, R125.
Lawton-Rauth, A. L., Alvarez-Buylla, E. R., Purugganan, M. D. (2000). Molecular Evolution of flower development. *Trends in Ecology and Evolution* **15**, 144-149.
Lerman, D. N., Michalak, P., Helin, A. B., Bettencourt, B. R. and Feder, M. E. (2003). Modification of heat-shock gene expression in Drosophila melanogaster populations via transposable elements. *Mol Biol Evol* **20**, 135-44.
Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-70.
Liang, H., Lin, Y. S. and Li, W. H. (2008). Fast evolution of core promoters in primate genomes. *Mol Biol Evol* **25**, 1239-44.
Lynch, M. (2007a). The evolution of genetic networks by non-adaptive processes. *Nat Rev Genet* **8**, 803-13.
Lynch, M. (2007b). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci U S A* **104 Suppl 1**, 8597-604.
Maddison, W. P. and Maddison, D. R. (2009). Mesquite: a modular system for evolutionary analysis, (ed.

Marin, O., Smeets, W. J. and Gonzalez, A. (1998). Evolution of the basal ganglia in tetrapods: a new perspective based on recent studies in amphibians. *Trends Neurosci* **21**, 487-94.
Martin, D. L. and Rimvall, K. (1993). Regulation of gamma-aminobutyric acid synthesis in the brain. *J Neurochem* **60**, 395-407.
Martyniuk, C. J., Awad, R., Hurley, R., Finger, T. E. and Trudeau, V. L. (2007). Glutamic acid decarboxylase 65, 67, and GABA-transaminase mRNA expression and total enzyme activity in the goldfish (Carassius auratus) brain. *Brain Res* **1147**, 154-66.
Matsuo, Y. and Yamazaki, T. (1984). Genetic Analysis of Natural Populations of DROSOPHILA MELANOGASTER in Japan. IV. Natural Selection on the Inducibility, but Not on the Structural Genes, of Amylase Loci. *Genetics* **108**, 879-896.
McMahon, A. P., Ingham, P. W. and Tabin, C. J. (2003). Developmental roles and clinical significance of hedgehog signaling. *Curr Top Dev Biol* **53**, 1-114.
Medina, L. and Reiner, A. (1995). Neurotransmitter organization and connectivity of the basal ganglia in vertebrates: implications for the evolution of basal ganglia. *Brain Behav Evol* **46**, 235-58.
Milunsky, A., Jick, H., Jick, S. S., Bruell, C. L., MacLaughlin, D. S., Rothman, K. J. and Willett, W. (1989). Multivitamin/folic acid supplementation in early pregnancy reduces the prevalence of neural tube defects. *Jama* **262**, 2847-52.
Montelione, G. T., Arrowsmith, C., Girvin, M. E., Kennedy, M. A., Markley, J. L., Powers, R., Prestegard, J. H. and Szyperski, T. (2009). Unique opportunities for NMR methods in structural genomics. *J Struct Funct Genomics* **10**, 101-6.
Mueller, T., Vernier, P. and Wullimann, M. F. (2006). A phylotypic stage in vertebrate brain development: GABA cell patterns in zebrafish compared with mouse. *J Comp Neurol* **494**, 620-34.
Muller, W. P. (1977). Diplotene chromosomes of Xenopus hybrid oocytes. *Chromosoma* **59**, 273-82.
Nair, R., Liu, J., Soong, T. T., Acton, T. B., Everett, J. K., Kouranov, A., Fiser, A., Godzik, A., Jaroszewski, L., Orengo, C. et al. (2009). Structural genomics is the largest contributor of novel structural leverage. *J Struct Funct Genomics* **10**, 181-91.
Nielsen, R. (2000). Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* **154**, 931-42.
Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., Fledel-Alon, A., Tanenbaum, D. M., Civello, D., White, T. J. et al. (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol* **3**, e170.
Notredame, C., Higgins, D. G. and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-17.
Oda-Ishii, I., Bertrand, V., Matsuo, I., Lemaire, P. and Saiga, H. (2005). Making very similar embryos with divergent genomes: conservation of regulatory mechanisms of Otx between the ascidians Halocynthia roretzi and Ciona intestinalis. *Development* **132**, 1663-74.
Ogino, H. and Ochi, H. (2009). Resources and transgenesis techniques for functional genomics in Xenopus. *Dev Growth Differ* **51**, 387-401.

Owens, D. F. and Kriegstein, A. R. (2002). Is there more to GABA than synaptic inhibition? *Nat Rev Neurosci* **3**, 715-27.
Pastinen, T., Sladek, R., Gurd, S., Sammak, A., Ge, B., Lepage, P., Lavergne, K., Villeneuve, A., Gaudin, T., Brandstrom, H. et al. (2004). A survey of genetic and epigenetic variation affecting human gene expression. *Physiol Genomics* **16**, 184-93.
Pinal, C. S. and Tobin, A. J. (1998). Uniqueness and redundancy in GABA production. *Perspect Dev Neurobiol* **5**, 109-18.
Popp, A., Urbach, A., Witte, O. W. and Frahm, C. (2009). Adult and embryonic GAD transcripts are spatiotemporally regulated during postnatal development in the rat brain. *PLoS ONE* **4**, e4371.
Posada, D. (2002). Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol Biol Evol* **19**, 708-17.
Raff, R. A., Kaufman, T.C. (1983). Embyos, genes, and evolution: the developmental-genetic basis of evolutionary change. Macmillan, New York.
Reiner, A., Medina, L. and Veenman, C. L. (1998). Structural and functional evolution of the basal ganglia in vertebrates. *Brain Res Brain Res Rev* **28**, 235-85.
Rockman, M. V. and Wray, G. A. (2002). Abundant raw material for cis-regulatory evolution in humans. *Mol Biol Evol* **19**, 1991-2004.
Ronald, J., Brem, R. B., Whittle, J. and Kruglyak, L. (2005). Local regulatory variation in Saccharomyces cerevisiae. *PLoS Genet* **1**, e25.
Rosenberg, N. A. and Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet* **3**, 380-90.
Sammut, B., Marcuz, A. and Pasquier, L. D. (2002). The fate of duplicated major histocompatibility complex class Ia genes in a dodecaploid amphibian, Xenopus ruwenzoriensis. *Eur J Immunol* **32**, 2698-709.
Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V. and Furlong, E. E. (2007). A core transcriptional network for early mesoderm development in Drosophila melanogaster. *Genes Dev* **21**, 436-49.
Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., Ruff, T. G., Milligan, S. B., Lamb, J. R., Cavet, G. et al. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297-302.
Schmid, M. and Steinlein, C. (1991). Chromosome banding in Amphibia. XVI. High-resolution replication banding patterns in Xenopus laevis. *Chromosoma* **101**, 123-32.
Segal, J. A., Barnett, J. L. and Crawford, D. L. (1999). Functional analyses of natural variation in Sp1 binding sites of a TATA-less promoter. *J Mol Evol* **49**, 736-49.
Smith, S. J., Fairclough, L., Latinkic, B. V., Sparrow, D. B. and Mohun, T. J. (2006). Xenopus laevis transgenesis by sperm nuclear injection. *Nat Protoc* **1**, 2195-203.
Somogyi, R., Wen, X., Ma, W. and Barker, J. L. (1995). Developmental kinetics of GAD family mRNAs parallel neurogenesis in the rat spinal cord. *J Neurosci* **15**, 2575-91.
Stark, A., Lin, M. F., Kheradpour, P., Pedersen, J. S., Parts, L., Carlson, J. W., Crosby, M. A., Rasmussen, M. D., Roy, S., Deoras, A. N. et al. (2007). Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. *Nature* **450**, 219-32.
Stern, D. L. (2000). Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079-91.

Streelman, J. T. and Kocher, T. D. (2002). Microsatellite variation associated with prolactin expression and growth of salt-challenged tilapia. *Physiol Genomics* **9**, 1-4.
Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437-60.
Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-95.
Taylor, J. S., Braasch, I., Frickey, T., Meyer, A. and Van de Peer, Y. (2003). Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res* **13**, 382-90.
Trabucchi, M., Trudeau, V. L., Drouin, G., Tostivint, H., Ihrmann, I., Vallarino, M. and Vaudry, H. (2008). Molecular characterization and comparative localization of the mRNAs encoding two glutamic acid decarboxylases (GAD65 and GAD67) in the brain of the African lungfish, Protopterus annectens. *J Comp Neurol* **506**, 979-88.
Tung, J., Fedrigo, O., Haygood, R., Mukherjee, S. and Wray, G. A. (2009). Genomic features that predict allelic imbalance in humans suggest patterns of constraint on gene expression variation. *Mol Biol Evol* **8**, 8.
Tymowska, J. (1991). Polyploidy and cytogenetic variation in frogs of the genus Xenopus. San Diego: Academic Press.
Waagepetersen, H. S., Sonnewald, U. and Schousboe, A. (1999). The GABA paradox: multiple roles as metabolite, neurotransmitter, and neurodifferentiative agent. *J Neurochem* **73**, 1335-42.
Wall, J. D. and Hudson, R. R. (2001). Coalescent simulations and statistical tests of neutrality. *Mol Biol Evol* **18**, 1134-5; author reply 1136-8.
Walser, J. C., Chen, B. and Feder, M. E. (2006). Heat-shock promoters: targets for evolution by P transposable elements in Drosophila. *PLoS Genet* **2**, e165.
Walters, J., Binkley, E., Haygood, R. and Romano, L. A. (2008). Evolutionary analysis of the cis-regulatory region of the spicule matrix gene SM50 in strongylocentrotid sea urchins. *Dev Biol* **315**, 567-78.
Watanabe, M., Maemura, K., Kanbara, K., Tamayama, T. and Hayasaki, H. (2002). GABA and GABA receptors in the central nervous system and other organs. *Int Rev Cytol* **213**, 1-47.
Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**, 256-76.
Wei, J. and Wu, J. Y. (2008). Post-translational regulation of L-glutamic acid decarboxylase in the brain. *Neurochem Res* **33**, 1459-65.
Wittkopp, P. J., Haerum, B. K. and Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85-8.
Wray, G. A. and Babbitt, C. C. (2008). Genetics. Enhancing gene regulation. *Science* **321**, 1300-1.
Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V. and Romano, L. A. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**, 1377-419.
Wright, S. (1931). Evolution in Mendelian Populations. *Genetics* **16**, 97-159.
Wullimann, M. F. and Mueller, T. (2004). Teleostean and mammalian forebrains contrasted: Evidence from genes to behavior. *J Comp Neurol* **475**, 143-62.

Zhang, X. and Borevitz, J. O. (2009). Global Analysis of Allele Specific Expression in Arabidopsis thaliana. *Genetics* **27**, 27.