THE MEIOTIC DRIVE MECHANISM OF A SELFISH CHROMOSOME IN ZEA MAYS

by

ELIZABETH GRACE LOWRY

(Under the Direction of R. Kelly Dawe)

ABSTRACT

Abnormal chromosome 10 (Ab10) is a selfish chromosome in maize. By exploiting the asymmetric nature of female meiosis, Ab10 eschews Mendel's law of equal segregation and passes into 80% of offspring in a single generation. This meiotic drive is caused by "neocentromere" movement in which knobs, or dense heterochromatic repeats, race along the meiotic spindle into the predestined seed. Knobs are composed of two distinct sequence repeats controlled by separate loci: Knob 180 and TR-1. Knob 180 neocentromeres correlate with meiotic drive and are activated by trans-acting loci located on the "distal tip" of Ab10, a euchromatic region that does not pair with any other region of the maize genome.

Ab10 was first discovered in 1942, but for 70 years the molecular mechanism of drive remained elusive. Here we present a comparative meiotic transcriptome analysis that identifies 9 genes unique to the distal tip of Ab10. Probing a BAC library homozygous for Ab10 allowed sequencing and assembly of corresponding genomic regions, resulting in construction of the first genetic map of the Ab10-specific genome. Two of five *suppressor of meiotic drive (smd*) Ab10 mutants, *smd3* and *smd8*, are characterized as large terminal deletions.

One of the nine distal genes is a C-terminal kinesin we call Kin618. Kin618 is a member of a multicopy gene family completely unique to the abnormal haplotype with closest homology

to native maize Kinesin 11. Identification of Ab10 mutant *smd12* as an epimutation, in which all copies of Kin618 have been silenced by DNA methylation, presents compelling evidence that Kin618 causes the drive phenotype by specifically moving Knob 180 neocentromeres. A second transcriptome analysis across haplotypes that vary in TR-1 neocentromeres did not identify any TR-1 specific factors. Contrary to hypothesis, TR-1 knobs are not targeted by a second highly evolved kinesin. In sum, the research presented here offers an illuminating look at the molecular mechanism of meiotic drive. By co-opting a native meiotic kinesin, the selfish chromosome evolved into a highly effective driver whose story plays out across the entire maize genome.

INDEX WORDS: meiotic drive, selfish gene, abnormal chromosome 10, epimutation, kinesin

THE MEIOTIC DRIVE MECHANISM OF A SELFISH CHROMOSOME IN ZEA MAYS

by

ELIZABETH GRACE LOWRY

BS, University of Virginia, 2009

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2015

© 2015

Elizabeth Grace Lowry

All Rights Reserved

THE MEIOTIC DRIVE MECHANISM OF A SELFISH CHROMOSOME IN ZEA MAYS

by

ELIZABETH GRACE LOWRY

Major Professor:	R. Kelly Dawe
Committee:	David Hall
	Wolfgang Lukowitz
	C. J. Tsai
	Jim Leebens-Mack

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
May 2015

ACKNOWLEDGEMENTS

I would like to thank my advisor, Kelly Dawe, as well as each member of my committee, Dave Hall, Wolfgang Lukowitz, C.J. Tsai, and Jim Leebens-Mack, for intellectual advice and experimental guidance while completing this research project. I would also like to thank my family, friends, and loved ones for support along the way.

This work was funded by fellowships from both the Graduate School of the University of Georgia and the Department of Genetics. Funding was also provided by a Doctoral Dissertation Improvement Grant from the National Science Foundation.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

Meiotic Drive and Ab10

In a paradox typical of science, the elegant pattern of Mendelian genetics stems from an essentially messy process. Accurate segregation of chromosomes during gamete formation, or meiosis, is accomplished by a complicated mechanism where mistakes carry a high cost. Meiosis is crucial, beautiful, ancient and yet strangely, fair. Strange especially when we consider the asymmetry inherent in female meiosis: of the four haploid products, only one develops into the egg or seed. Some genes (or groups of genes, or entire chromosomes) seem to know this, and will go to great lengths to be included in that predestined cell. If genes are especially clever, they break meiotic rules and show up in the egg, and therefore the offspring, more than 50% of the time. These genes are known as selfish elements [1].

Selfishness is an anthropomorphizing, though accurate, metaphor. Selfish elements are often genetically linked to deleterious alleles, creating strong evolutionary pressure to select for Mendelian segregation [2,3]. If selfish elements break the rules of meiosis to preferentially segregate, the process is referred to as meiotic drive [4]. Most of the selfish elements we know about do not exhibit true meiotic drive. Segregation levels greater than 50% are seen in the t-haplotype of mouse and the Drosophila sex-ratio (SR) and segregation distorter (SD) systems, but this preferential segregation happens after meiosis [5–7]. The most convincing evidence of meiotic drive comes from the abnormal chromosome 10 (Ab10) system in maize.

Ab10 is a haplotype variant of normal chromosome 10 (N10) that occurs across natural maize populations at levels approaching 20% [8]. Ab10 differs from normal chromosome 10 on the long arm of the chromosome that is serendipitously linked to the R marker, which turns the pericarp of corn kernels purple. The drive of Ab10 can be noticed in a single cross, in which 80% of the kernels are purple [9]. Ab10 is also distinguished from N10 by the presence of "knobs," large areas of tandem repeats that bunch together into balled heterochromatin. There are two distinct sequences which compose knobs: a 180 base pair sequence known as Knob 180 and a 350 base pair sequence known as TR-1 [10,11]. Knobs are not only found on the abnormal chromosome, but are present at 33 distinct locations throughout the maize genome [12,13]. Knobs are quiescent in the absence of Ab10, but when Ab10 is present in the plant all knobs and genetically linked loci preferentially segregate.

Marcus Rhoades, a maize cytologist struck by the unusual behavior of Ab10 knobs during meiosis, first characterized the abnormal chromosome. In canonical chromosome division, chromosomes attach via their centromere to the proteinaceous kinetochore and are then pulled to opposite ends of the dividing cell by a combination of microtubule flux and kinesins, motor proteins that travel along microtubules [14]. When Ab10 is present in the plant, knobs attach to the spindle and streak laterally along the microtubules towards the poles of the dividing cell, outracing the centromere [15,16]. Rhoades reasoned that this "neocentromere" behavior of knobs causes the preferential segregation of Ab10 in female meiosis. In maize, the four products of megasporogenesis are arranged vertically in the plant and it is always the bottom cell that is fertilized by the pollen and develops into the seed. By establishing and maintaining a polar orientation, Ab10 targets this bottom-most cell and therefore reaches segregation levels above 50%.

As neocentromere movement of all genomic knobs is only observed while Ab10 is present, the genes that cause preferential segregation must be trans-acting and located in sequence unique to the abnormal chromosome [17]. The long arm of Ab10 is distinguished from N10 by three small TR-1 knobs, or chromomeres, followed distally by a "differential region" containing genes shared with N10 but subjected to a series of multiple inversions that prevent pairing and recombination between the two chromosomes [18]. Proximal to the differential region is a large Knob 180 and past that is the "distal tip," a region of euchromatin that appears extrachromosomal, containing no other genes found in the maize genome. Knob 180 and TR-1 neocentromeres are controlled by genetically distinct loci, with TR-1 activation mapped near the three small TR-1 knobs and Knob 180 movement mapped to the euchromatic distal tip [17]. The genes that cause preferential segregation and meiotic drive also map to the distal tip, in the same region as Knob 180 movement. This mapping was performed using a series of deficiency lines that contain truncated Ab10 chromosomes with breakpoints arrayed along the long arm [17]. The smallest truncation, Df(L), contains all of the Abnormal Chromosome save the distal tip and does not have Knob 180 neocentromeres nor meiotic drive [19,20]. By contrast, Df(L) still has TR-1 neocentromeres, clearly demarcating the functional difference between the two knob types.

To date, there are four different haplotype variants of the 10th chromosome that vary in knob content and drive ability [21]. Ab10-I was first characterized by Rhoades and is described above. Ab10-II has only one small TR-1 knob and two large Knob 180 knobs. Ab10-II does not initiate TR-1 neocentromeres, yet it segregates at levels similar to Ab10-I. Ab10-III is similar in knob content to Ab10-I except instead of a large Knob 180, it has a mixed knob containing TR-1 repeats as well. Ab10-III also preferentially segregates, though its neocentromere activity profile has yet to be characterized. The fourth variant, K10L2, has only TR-1 knobs and basically

negligible levels of drive, measured at a statistically significant 51%. K10L2 causes the formation of TR-1 neocentromeres but not Knob 180 neocentromeres. Interestingly, when a plant is heterozygous for either Ab10-I or Ab10-II and K10L2, the drive of Ab10-I or Ab10-II is reduced [21]. This data prompts the theory that K10L2 suppresses drive of the Abnormal chromosomes and that TR-1 and Knob 180 are locked in an arms-race, using separate loci to drive to greater propagation within the maize genome.

The final evidence for the functional distinction between TR-1 and Knob 180 repeats comes from a suite of Ab10-I mutants. These suppressor of meiotic drive (*smd*) mutants were created using Robertson's Mutator (*Mu*) mutagenesis. In total, five mutants were isolated with greatly reduced or negligible levels of drive: *smd1*, *smd3*, *smd8*, *smd12*, and *smd13* [19,22]. All five mutants share the phenotype of a loss of Knob 180 neocentromeres. However, all five mutants still activate TR-1 neocentromeres. Like the deficiency line Df(L) and the chromosome 10 haplotype variants, the mutants confirm the link between Knob 180 neocentromere movement and meiotic drive.

Kinesins and Chromosome Movement

The unique style of neocentromere movement suggests the involvement of a motor protein that allows lateral sliding along microtubules towards the poles of the dividing meiotic cell. Kinesins are molecular motor proteins that hydrolyze ATP to move along the bipolar spindle in either the plus or minus direction. Kinesins are characterized by an approximately 350 amino acid "motor domain" composed of the catalytic core, responsible for hydrolyzing ATP, a binding site for microtubules, and the "neck" which determines the directionality of movement [23]. The position of the motor domain within the protein sequence confers directionality of

movement: “C-type” kinesins have a motor domain at the C-terminus and move towards the minus end of microtubules while “N-type” kinesins have a motor domain at the N-terminus and moves towards the plus end of microtubules. A third variant, the “I-type,” lacks a neck and has the motor domain located in the center of the amino acid sequence. Characterized I-type kinesins move to both the plus and minus ends of the spindle [24,25].

The basic form of kinesins is conserved across kingdoms. The “tail” or “stalk” domain is composed of a α-helical coiled coil and is highly divergent in sequence and length, as it is specific to the function of the individual kinesin. The tail of the kinesin binds to the cargo, which may be proteins, organelles, RNA, microtubules, or another kinesin subunit if the kinesin functions as a holoenzyme. Kinesin holoenzymes may be either homo or hetero monomers, dimers, trimers, or tetramers. The tail domain also may bind kinesin associated proteins, or KAPs. KAPs may work to target kinesins to different organelles [26]. Phylogenetic analyses based on the full kinesin sequence have divided all characterized kinesins into 14 families [23,24]. Families vary both by directionality of movement and function which include transporting organelles, signal transduction, and moving chromosomes and spindles during meiosis and mitosis.

As a testament to the importance of kinesins in chromosome movement, six of the fourteen families of kinesins have members associated with cell division. The Kinesin-4 family, previously known as “chromokinesins,” is notable in that its members bind directly to chromosomes. KIF4, originally called Chromokinesin, contains a basic-leucine zipper DNA-binding domain that directly contacts DNA in vitro [27]. Later work shows that KIF4 interacts with condensin to promote compaction of chromosome arms during mitosis [28]. KIF4 also functions in the formation of the central spindle and midbody at the end of meiosis [29–31]. Members of the

Kinesin-5 family, which function as homotetramers, have been shown through mutational analysis to be crucial in the separation of the spindle pole bodies during mitosis in *Aspergillus* [32]. The Kinesin-6 family contains Rab6-KIFL, which is necessary for cytokinesis [33]. The Kinesin-7 family contains CENP-E, a member of the proteinaceous kinetochore involved in MT-chromosome attachment and chromosomal alignment at metaphase [34–37]. The Kinesin-8 family includes Klp5 and Klp6, which in S. *Pombe* are necessary for proper sister chromatid separation [38].

The Kinesin-10 family, notable for a helix-hairpin-helix DNA binding motif in the tail region, has many well-characterized members such as KID, Kif22, and NOD, that all play prominent roles in cell division. Human KID kinesin is seen enriched on the kinetochore during anaphase and colocalizes with mitotic chromosomes [39]. KID has also been found to be involved in microtubule bundling activity necessary for maintaining metaphase spindle size [40]. Work on XKid, the *Xenopus* homologue of KID, suggests that the kinesin helps align chromosomes at metaphase by pushing chromosome arms towards the equator of the metaphase spindle until the start of anaphase [41]. It may be the HhH DNA binding motif that binds directly to chromosome arms for this function, though that has yet to be directly proven. The KID homolog in mouse, Kif22, associates with the mitotic checkpoint protein Chfr which in turn regulates chromosomal stability [42]. NOD, in *Drosophila melanogaster*, localizes to chromosome arms during meiosis and is necessary for achiasmate chromosome segregation [43,44]. Like XKid, NOD facilitates metaphase alignment by pushing chromosomes arms toward the metaphase plate, perhaps by promoting tubulin polymerization at the plus-end of microtubules [45,46]. Interestingly, the C-terminus tail of NOD contains two DNA binding domains. The first spans about 70 amino acids and contains a high-mobility group N (HMGN) motif necessary for localization to chromosome

arms [44,47]. The second region is a helix-hairpin-helix(2)/Nod-like domain, a non-specific motif which only localizes to chromosomes during the oocyte stage and likely functions in ways other than DNA binding, such as mediating protein-protein interactions [48,49].

Five of the six kinesin families involved in chromosome segregation have N-terminal motor domains and move towards the plus-end of the microtubule spindle. By contrast, the Kinesin-14 family is characterized a motor-domain at the C-terminus and moves towards the minus-end of microtubules. Within the Kinesin-14 family there are two subfamilies distinguished by their conserved "neck" residues. The Kinesin-14A subfamily contains members involved in chromosome segregation including mouse KIFC1, hamster CHO2, human HSET, *Drosphila* Ncd, yeast Kar3, and *Arabidopsis* ATK1. These kinesins promote the organization and stability of the meiotic and mitotic spindles by binding to tubulin in both their tail and motor domains and either bundling microtubules or promoting microtubule-microtubule sliding [50,51]. This activity regulates spindle length in animals and focuses spindle poles in plants. *Drosophila* Ncd is responsible for the fidelity of division in both meiosis and mitosis, with its absence leading to abnormally wide spindles [52]. Yeast Kar3 regulates spindle length by associating with one of two non-motor polypeptides, Cik1 or Vik1, and forming a heterodimer [53,54]. The loss of *Arabidopsis* ATK1 results in abnormal spindle organization with multi-axial poles at Metaphase I, making it necessary for male fertility [55]. Members of this family may regulate spindle fidelity through microtubule plus-end tracking, a function that has been observed in both Ncd and HSET [56,57]. Interestingly, there is evidence that some family members may be multi-copy. In mouse, kinesins KIFC1 and KIFC5, originally thought to be two distinct kinesins, differ only in 2 small sequence insertions in the tail and a number of SNPs in the motor domain [58]. Further bioinformatic analysis lead to the hypothesis that the two are members of a multi-gene family mapping to four

distinct genetic loci that generate alternatively spliced cDNA species slightly divergent in their tail domain [58]. As the proper segregation of chromosomes is crucial for organismal success, it is unsurprising that the genome contains multiple copies of particular kinesins.

DNA Methylation and Epimutants in Plants

During Prophase, Ab10 "knobs" bunch into heterochromatic clusters. Heterochromatic DNA in plants is characterized by an excess of repeats, transposable elements (TEs), and low genic transcriptional activity [59]. Biochemically, plant heterochromatin is recognized by a series of epigenetic modifications including histone H3 lysine 9 dimethylation (H3K9me2) and cytosine methylation occurring at all three possible sequence motifs: CG, CHG, and CHH [60,61]. Though all types of DNA methylation are enriched in pericentromeric regions, characterized by high levels of repeats and TEs, transcribed genes in *Arabidopsis* have a singular pattern of CG methylation. CG methylation, usually a sign of repressed chromatin, is present at high levels across gene bodies. However, a sharp dip in CG methylation in the promoter presumably makes the sequence available to bind polymerase and be transcribed [62,63]. By contrast, CHG and CHH methylation is more likely found in repetitive regions or in transposons [63]. Plant de novo cytosine methylation occurs the same way for all three sequence motifs via the RNA-directed DNA methylation (RdDM) pathway, mediated by the DNA methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) [64]. After DNA replication, the maintenance methylation pathway depends on the sequence context. CHH DNA methylation is maintained through the RdDM pathway. CG methylation is maintained by DNA METHYLTRANSFERASE 1 (MET1) and CHG methylation is maintained by CHROMOMETHYLASE 3 (CMT3), which is guided in turn by the H3K9me2 mark [65,66].

DNA methylation patterns provide another layer of genetic variation that can contribute to phenotypic diversity. Epigenetic changes may lead to epialleles, where a change in methylation has altered the transcription of a gene and caused a phenotype [67]. These epimutants, as they are also called, have been found to alter floral morphology, gene expression, and disease resistance in *Arabidopsis thaliana*, pigmentation in *Zea mays*, and floral symmetry in *Linaria vulgaris* [68–72]. Hypermethylation of the coding region of the SUPERMAN gene in *Arabidopsis*, effecting floral morphology, leads to a phenotype similar to loss-of-function mutations, showing that methylation changes can essentially turn off a gene [68]. Hypermethylation of the Lcyc gene, controlling floral symmetry in toadflax (*Linaria vulgaris*), also leads to complete transcriptional silencing in the mutant [72].

The insertion of transposable elements may also cause epimutants. In *Sinapsis alba* (yellow mustard), the insertion of a Copia-like retrotransposon in the 5′ untranslated region leads to cytosine methylation in the promoter region of the FATTY ACID ELONGATION (FAE1) gene, reducing expression of the gene [73]. In *Zea mays*, the tandemly repeated multicopy allele of pericarp color1 (p1) is known to vary phenotypically depending on the methylation state of its alleles [71]. Mutagenesis with the DNA transposon Robertson's Mutator (*Mu*) produced a p1 allele with an inserted *Mu* that caused hypomethylation of a floral enhancer 4.7 kb upstream of the insertion site. This epimutation resulted in expression from gene copies that are otherwise suppressed [74]. Though evidence from several plant species points to a close relationship between transposons, epigenetic change, and epialleles, the exact nature of the interaction remains to be characterized.

It's long been known that epigenetic changes control the expression of transposable elements, leading some to theorize DNA methylation in plants evolved as a defense mechanism

against TEs [75]. In *Arabidopsis*, the DNA transposon Robertson's Mutator (*Mu*) has been found to be regulated by the chromatin remodeling gene Decrease in DNA Methylation (DDM1) [76]. Loss of DDM1 leads to upregulation of *Mu*. The same results are seen in *Zea mays*, with the loss of RNA-dependent RNA polymerase 2 (RDR2), a component of the RdDM pathway, leading to widespread reactivation of transcriptionally silenced *Mu* transposons [77]. The characterization of a naturally occurring Mu killer (*Muk*) locus, also in *Zea mays*, which specifically silences only *Mu* transposons, may represent an adaptive defense of the plant to keep *Mu* in check [78]. *Muk* arose from an inverted duplication of a truncated *Mutator* transposon, encoding a hairpin transcript that silences all genomic *Mu* elements [79]. The origin of *Muk* demonstrates how a small DNA sequence change can have epigenetic consequences that affect the entire genome.

Purpose of Study

As a scientific community, we know surprisingly little about the mechanics of meiotic drive. Though maize abnormal chromosome 10 has been studied since its discovery in 1942, we still have no molecular evidence supporting the theory of Ab10 preferential segregation. The key to understanding drive lies in the mechanism of Knob 180 neocentromere movement. The genes that control both drive and Knob 180 neocentromeres lie on the "distal tip" of the abnormal chromosome, a region that is the focus of this dissertation. Non-centromeric meiotic chromosome movement is not well understood and characterizing the Ab10 driver will also contribute to our understanding of chromosome motility during meiosis.

References

1. Dawkins, R. *The Selfish Gene*. (Oxford University Press: 1989).at <http://books.google.com/books?id=WkHO9HI7koEC>
2. Crow, J. F. Why is Mendelian segregation so exact? *BioEssays: news and reviews in molecular, cellular and developmental biology* **13**, 305–12 (1991).
3. Leigh, E. G. How does selection reconcile individual advantage with the good of the group? *Proceedings of the National Academy of Sciences of the United States of America* **74**, 4542–6 (1977).
4. BURT, A., Trivers, R. & Burt, A. *Genes in Conflict: The Biology of Selfish Genetic Elements*. (Harvard University Press: 2009).at <http://books.google.com/books?id=8e8xLbmtuMcC>
5. Crismani, W., Girard, C. & Mercier, R. Tinkering with meiosis. *Journal of experimental botany* **64**, 55–65 (2013).
6. Larracuente, A. M. & Presgraves, D. C. The selfish Segregation Distorter gene complex of Drosophila melanogaster. *Genetics* **192**, 33–53 (2012).
7. Manser, A., Lindholm, A. K., König, B. & Bagheri, H. C. The effect of polyandry on a distorter system with differential viabilities in the sexes. *Communicative & integrative biology* **5**, 550–2 (2012).
8. McClintock, Yamakake T, B. A. Chromosome Constitution of Races of Maize: its significance in the interpretation of relationships between races and varieties in the Americas. *Colegio de Postgraduados: Chapingo, Mexico* (1981).
9. Rhoades, M. M. Preferential Segregation in Maize. *Genetics* **27**, 395–407 (1942).
10. Ananiev, E. V, Phillips, R. L. & Rines, H. W. A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10785–90 (1998).
11. Dennis, E. S. & Peacock, W. J. Knob heterochromatin homology in maize and its relatives. *Journal of molecular evolution* **20**, 341–50 (1984).
12. Albert, P. S., Gao, Z., Danilova, T. V & Birchler, J. A. Diversity of chromosomal karyotypes in maize and its relatives. *Cytogenetic and genome research* **129**, 6–16 (2010).
13. Kato, Y. Cytological studies of maize (Zea mays L.) and teosinte (Zea mexicana Shrader Kuntze) in relation to their orIgin and evolution. *Mass Agric Exp Sta Bull* **635**, 1–185 (1976).
14. Cheeseman, I. M. & Desai, A. Molecular architecture of the kinetochore-microtubule interface. *Nature reviews. Molecular cell biology* **9**, 33–46 (2008).
15. Rhoades, M. Preferential segregation in maize. *Heterosis* 66–80 (1952).
16. Yu, H. G., Hiatt, E. N., Chan, A., Sweeney, M. & Dawe, R. K. Neocentromere-mediated chromosome movement in maize. *The Journal of cell biology* **139**, 831–40 (1997).
17. Hiatt, E. N., Kentner, E. K. & Dawe, R. K. Independently regulated neocentromere activity of two classes of tandem repeat arrays. *The Plant cell* **14**, 407–20 (2002).
18. Mroczek, R. J., Melo, J. R., Luce, A. C., Hiatt, E. N. & Dawe, R. K. The maize Ab10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. *Genetics* **174**, 145–54 (2006).
19. Hiatt, E. N. & Dawe, R. K. Four loci on abnormal chromosome 10 contribute to meiotic drive in maize. *Genetics* **164**, 699–709 (2003).

20. Kanizay, L. B. The variants of maize chromosome 10 and their roles in meiotic drive. 111 (2011).
21. Kanizay, L. B., Albert, P. S., Birchler, J. A. & Dawe, R. K. Intragenomic conflict between the two major knob repeats of maize. *Genetics* **194**, 81–9 (2013).
22. Dawe, R. K. & Cande, W. Z. Induction of centromeric activity in maize by suppressor of meiotic drive 1. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8512–7 (1996).
23. Lawrence, C. J. *et al.* A standardized kinesin nomenclature. *The Journal of cell biology* **167**, 19–22 (2004).
24. Miki, H., Okada, Y. & Hirokawa, N. Analysis of the kinesin superfamily: insights into structure and function. *Trends in cell biology* **15**, 467–76 (2005).
25. Lawrence, C. J., Malmberg, R. L., Muszynski, M. G. & Dawe, R. K. Maximum likelihood methods reveal conservation of function among closely related kinesin families. *Journal of molecular evolution* **54**, 42–53 (2002).
26. Gyoeva, F. K., Bybikova, E. M. & Minin, A. A. An isoform of kinesin light chain specific for the Golgi complex. *Journal of cell science* **113 (Pt 1**, 2047–54 (2000).
27. Wang, S. Z. & Adler, R. Chromokinesin: a DNA-binding, kinesin-like nuclear protein. *The Journal of cell biology* **128**, 761–8 (1995).
28. Samejima, K. *et al.* Mitotic chromosomes are compacted laterally by KIF4 and condensin and axially by topoisomerase IIα. *The Journal of cell biology* **199**, 755–70 (2012).
29. Kurasawa, Y., Earnshaw, W. C., Mochizuki, Y., Dohmae, N. & Todokoro, K. Essential roles of KIF4 and its binding partner PRC1 in organized central spindle midzone formation. *The EMBO journal* **23**, 3237–48 (2004).
30. Castoldi, M. Chromokinesin Xklp1 Contributes to the Regulation of Microtubule Density and Organization during Spindle Assembly. *Molecular Biology of the Cell* **17**, 1451–1460 (2005).
31. Hu, C.-K., Coughlin, M., Field, C. M. & Mitchison, T. J. KIF4 regulates midzone length during cytokinesis. *Current biology: CB* **21**, 815–24 (2011).
32. Enos, A. P. & Morris, N. R. Mutation of a gene that encodes a kinesin-like protein blocks nuclear division in A. nidulans. *Cell* **60**, 1019–27 (1990).
33. Hill, E., Clarke, M. & Barr, F. A. The Rab6-binding kinesin, Rab6-KIFL, is required for cytokinesis. *The EMBO journal* **19**, 5711–9 (2000).
34. Wood, K. W., Sakowicz, R., Goldstein, L. S. & Cleveland, D. W. CENP-E is a plus end-directed kinetochore motor required for metaphase chromosome alignment. *Cell* **91**, 357–66 (1997).
35. Schaar, B. T., Chan, G. K., Maddox, P., Salmon, E. D. & Yen, T. J. CENP-E function at kinetochores is essential for chromosome alignment. *The Journal of cell biology* **139**, 1373–82 (1997).
36. McEwen, B. F. *et al.* CENP-E is essential for reliable bioriented spindle attachment, but chromosome alignment can be achieved via redundant mechanisms in mammalian cells. *Molecular biology of the cell* **12**, 2776–89 (2001).
37. Sardar, H. S., Luczak, V. G., Lopez, M. M., Lister, B. C. & Gilbert, S. P. Mitotic kinesin CENP-E promotes microtubule plus-end elongation. *Current biology: CB* **20**, 1648–53 (2010).
38. West, R. R., Malmstrom, T. & McIntosh, J. R. Kinesins klp5+ and klp6+ are required for normal chromosome movement in mitosis. *J. Cell Sci.* **115**, 931–940 (2002).

39. Tokai, N. *et al.* Kid, a novel kinesin-like DNA binding protein, is localized to chromosomes and the mitotic spindle. *The EMBO journal* **15**, 457–67 (1996).
40. Tokai-Nishizumi, N., Ohsugi, M., Suzuki, E. & Yamamoto, T. The chromokinesin Kid is required for maintenance of proper metaphase spindle size. *Molecular biology of the cell* **16**, 5455–63 (2005).
41. Funabiki, H. & Murray, A. W. The Xenopus Chromokinesin Xkid Is Essential for Metaphase Chromosome Alignment and Must Be Degraded to Allow Anaphase Chromosome Movement. *Cell* **102**, 411–424 (2000).
42. Maddika, S., Sy, S. M.-H. & Chen, J. Functional interaction between Chfr and Kif22 controls genomic stability. *The Journal of biological chemistry* **284**, 12998–3003 (2009).
43. Zhang, P., Knowles, B. A., Goldstein, L. S. & Hawley, R. S. A kinesin-like protein required for distributive chromosome segregation in Drosophila. *Cell* **62**, 1053–62 (1990).
44. Afshar, K., Barton, N. R., Hawley, R. S. & Goldstein, L. S. DNA binding and meiotic chromosomal localization of the Drosophila nod kinesin-like protein. *Cell* **81**, 129–38 (1995).
45. Theurkauf, W. E. & Hawley, R. S. Meiotic spindle assembly in Drosophila females: behavior of nonexchange chromosomes and the effects of mutations in the nod kinesin-like protein. *The Journal of cell biology* **116**, 1167–80 (1992).
46. Cui, W. *et al.* Drosophila Nod protein binds preferentially to the plus ends of microtubules and promotes microtubule polymerization in vitro. *Molecular biology of the cell* **16**, 5400–9 (2005).
47. Afshar, K., Scholey, J. & Hawley, R. S. Identification of the chromosome localization domain of the Drosophila nod kinesin-like protein. *The Journal of cell biology* **131**, 833–43 (1995).
48. Cui, W. & Hawley, R. S. The HhH2/NDD domain of the Drosophila Nod chromokinesin-like protein is required for binding to chromosomes in the oocyte nucleus. *Genetics* **171**, 1823–35 (2005).
49. Doherty, A. J., Serpell, L. C. & Ponting, C. P. The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic acids research* **24**, 2488–97 (1996).
50. Fink, G. *et al.* The mitotic kinesin-14 Ncd drives directional microtubule-microtubule sliding. *Nature cell biology* **11**, 717–23 (2009).
51. Walczak, C. E., Verma, S. & Mitchison, T. J. XCTK2: a kinesin-related protein that promotes mitotic spindle assembly in Xenopus laevis egg extracts. *The Journal of cell biology* **136**, 859–70 (1997).
52. Hatsumi, M. & Endow, S. A. Mutants of the microtubule motor protein, nonclaret disjunctional, affect spindle structure and chromosome movement in meiosis and mitosis. *Journal of cell science* **101 (Pt 3**, 547–59 (1992).
53. Chu, H. M. A. *et al.* Kar3 interaction with Cik1 alters motor structure and function. *The EMBO journal* **24**, 3214–23 (2005).
54. Allingham, J. S., Sproul, L. R., Rayment, I. & Gilbert, S. P. Vik1 modulates microtubule-Kar3 interactions through a motor domain that lacks an active site. *Cell* **128**, 1161–72 (2007).
55. Chen, C. *et al.* The Arabidopsis ATK1 gene is required for spindle morphogenesis in male meiosis. *Development (Cambridge, England)* **129**, 2401–9 (2002).

56. Goshima, G., Wollman, R., Stuurman, N., Scholey, J. M. & Vale, R. D. Length control of the metaphase spindle. *Current biology: CB* **15**, 1979–88 (2005).
57. Braun, M. *et al.* The human kinesin-14 HSET tracks the tips of growing microtubules in vitro. *Cytoskeleton (Hoboken, N.J.)* **70**, 515–21 (2013).
58. Navolanic, P. M. Identification of Isoforms of a Mitotic Motor in Mammalian Spermatogenesis. *Biology of Reproduction* **62**, 1360–1369 (2000).
59. Lippman, Z. & Martienssen, R. The role of RNA interference in heterochromatic silencing. *Nature* **431**, 364–70 (2004).
60. Henderson, I. R. & Jacobsen, S. E. Epigenetic inheritance in plants. *Nature* **447**, 418–24 (2007).
61. Martienssen, R. A. & Colot, V. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science (New York, N.Y.)* **293**, 1070–4 (2001).
62. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**, 1189–201 (2006).
63. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**, 215–9 (2008).
64. Wassenegger, M., Heimes, S., Riedel, L. & Sänger, H. L. RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76**, 567–76 (1994).
65. Du, J. *et al.* Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**, 167–80 (2012).
66. Kankel, M. W. *et al.* Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* **163**, 1109–22 (2003).
67. Schmitz, R. J. *et al.* Transgenerational epigenetic instability is a source of novel methylation variants. *Science (New York, N.Y.)* **334**, 369–73 (2011).
68. Jacobsen, S. E. & Meyerowitz, E. M. Hypermethylated SUPERMAN epigenetic alleles in arabidopsis. *Science (New York, N.Y.)* **277**, 1100–3 (1997).
69. Bender, J. & Fink, G. R. Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. *Cell* **83**, 725–34 (1995).
70. Stokes, T. L., Kunkel, B. N. & Richards, E. J. Epigenetic variation in Arabidopsis disease resistance. *Genes & development* **16**, 171–82 (2002).
71. Cocciolone, S. M., Chopra, S., Flint-Garcia, S. A., McMullen, M. D. & Peterson, T. Tissue-specific patterns of a maize Myb transcription factor are epigenetically regulated. *The Plant journal: for cell and molecular biology* **27**, 467–78 (2001).
72. Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**, 157–61 (1999).
73. Zeng, F. & Cheng, B. Transposable Element Insertion and Epigenetic Modification Cause the Multiallelic Variation in the Expression of FAE1 in Sinapis alba. *The Plant Cell* **26**, 2648–2659 (2014).
74. Robbins, M. L., Sekhon, R. S., Meeley, R. & Chopra, S. A Mutator transposon insertion is associated with ectopic expression of a tandemly repeated multicopy Myb gene pericarp color1 of maize. *Genetics* **178**, 1859–74 (2008).
75. Matzke, M. A., Mette, M. F. & Matzke, A. J. Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. *Plant molecular biology* **43**, 401–15 (2000).

76. Singer, T., Yordan, C. & Martienssen, R. A. Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). *Genes & development* **15**, 591–602 (2001).
77. Jia, Y. *et al.* Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. *PLoS genetics* **5**, e1000737 (2009).
78. Slotkin, R. K., Freeling, M. & Lisch, D. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nature genetics* **37**, 641–4 (2005).
79. Slotkin, R. K., Freeling, M. & Lisch, D. Mu killer causes the heritable inactivation of the Mutator family of transposable elements in Zea mays. *Genetics* **165**, 781–97 (2003).

CHAPTER 2: DISCOVERY AND CHARACTERIZATION OF THE UNIQUE GENETIC CONTENT OF ABNORMAL CHROMOSOME 10 IN MAIZE

Abstract

Though meiotic drive of maize Abnormal Chromosome 10 (Ab10) was first discovered in 1942 by Marcus Rhoades, the molecular mechanism of drive remains a mystery. The trans-acting drive factors are located on the distal tip of Ab10, an area of extrachromosomal DNA that does not pair with any other chromosome in the genome. However, for 70 years the DNA content of the distal tip was unknown, preventing further characterization. Here we perform a comparative transcriptome analysis in which we use RNA-seq and de novo assembly to discover meiotic transcripts originating from the distal tip of Ab10. This approach identifies 9 distinct genes that map to the same region as the drive genes and are potential factors involved in meiotic drive. Using a combination of genomic BAC clones, homozygous deficiency lines, and a suite of Ab10 mutant lines, we have characterized the first genomic sequence both unique and specific to the abnormal haplotype and developed the first genetic map of the distal tip.

Introduction

Though segregation distortion has been identified in a variety of organisms, meiotic drive is a rare disruptive process that results in the violation of Mendel's First Law of Segregation,

specifically during meiosis[1]. The abnormal chromosome 10 (Ab10) haplotype in maize is a classic example of meiotic drive, in which heterochromatic knobs preferentially segregate into offspring with frequency levels between 70-75%, far above the 50% predicted by Mendel [2]. Knobs are large clusters of tandem repeats that ball up into heterochromatin, thus gaining their nickname. There are two sequences that form knobs: a 180-bp long sequence (Knob 180) and a 350-bp long sequence (TR-1) [3,4]. Ab10 itself has three small TR-1 knobs and a large Knob 180. Knobs are found on all nine other chromosomes, however most maize strains have only a few. In total, knobs have been identified at 34 locations within the genome of both maize and teosinte[5]. In the absence of the abnormal chromosome, knobs segregate normally during meiosis. When Ab10 is present, knobs are driven to preferential segregation as a result of trans-acting factors located on Ab10.

Drive only occurs during female meiosis, since only one cell of megasporogenesis is fertilized by pollen and develops into the seed. When Ab10 is present, knobs transform into "neocentromeres" and streak along the microtubule spindle during anaphase ahead of the rest of the dividing chromosomes. Neocentromere activity pulls the knobs into the top and bottom cells of the linear tetrad and it is the bottom cell that becomes the egg. Knobs may vary in their degree of preferential segregation depending on the efficiency of neocentromere formation and the likelihood of recombination between the knob and the centromere [6]. Though both Knob 180 and TR-1 display neocentromere activity when Ab10 is present, they are controlled by separate trans-acting genes [7].

Abnormal chromosome 10 is much longer than normal 10 with an "extrachromosomal" region that does not pair with any other part of the maize genome. The haplotype begins distal to the R1 gene, a commonly used genetic marker effecting kernel color often used to score the

presence of Ab10 (Figure 2.1). Just distal to R1 is a differential segment with the three TR-1 knobs and the gene(s) that causes TR-1 neocentromere activity. Past that is a shared region containing genes present on N10 (L13, O7, W2, SR2) that have been rearranged by multiple inversions [8]. This inverted segment prevents Ab10 from recombining with N10 and maintains linkage between the trans-acting factors that distort segregation and their targets (knobs) which is necessary in a drive system [9]. The inversions are followed by a large Knob 180 and the distal tip, a region of previously unknown sequence unique to the abnormal haplotype.

A series of Ab10 mutations and deficiency lines have been used to map the trans-acting factors involved in Knob 180 neocentromere formation and drive to the distal tip. One particularly useful deficiency line, Df(L), contains the entire abnormal chromosome except for the distal tip and does not meiotically drive nor form Knob 180 neocentromeres (Figure 2.1)[6,10]. Mutations of Ab10 have been identified in a forward mutation screen using the active transposable element Robertson's mutator (*Mu*). Mutants of Ab10 have normal (50%) segregation levels of the selfish chromosome and are called *suppressor of meiotic drive* (*smd* mutants). Five mutants have been identified in the Dawe lab. Two of those, *smd1* (*suppressor of meiotic drive 1*) and *smd3*, are published but all five share the same phenotype [6,11]. All five mutants (*smd1, smd3, smd8*, *smd12*, and *smd13*) are unable to preferentially segregate.

Here we use the Ab10 deficiency line Df(L) as well as the *smd* mutants to identify and characterize a suite of RNA transcripts originating from the distal tip of the abnormal haplotype. By finally obtaining genomic data unique to the selfish chromosome, we may better understand the meiotic drive haplotype in the context of the maize genome.

Materials and Methods

Comparative RNA-seq and de novo Transcriptome Assembly

To identify transcripts unique to Ab10, a comparison was made between heterozygous Ab10/+ plants and heterozygous Df(L)/+ plants. Both lines had been backrossed at least six times into Dawe lab inbred that is closely related to the W23 inbred (but carrying the dominant C1 allele). Also included in the bioinformatic assembly were Illumina reads from a homozygous Ab10/Ab10 line and a sibling N10/N10 line collected previously by Lisa Kanizay (Table 2.1). For the Ab10/+ and Df(L)/+ lines, plants were grown in a greenhouse until tassel formation at which time the anthers, containing meiocytes, were extruded into PBS and dissected to identify meiocytes at the correct stage of meiosis. Properly staged anthers were then collected into 1.5 mL eppendorf tubes and flash frozen with liquid nitrogen. RNA extraction from whole anthers was performed with the Qiagen RNeasy kit. From each of the two genotypes, RNA was produced from three plants and combined into one tube. Samples were sent to Georgia Genomics Facility (GGF) which created libraries using the Truseq RNA Sample Preparation kit from Illumina. The libraries were then barcoded, pooled and analyzed in a single SE100 lane outsourced to the University of Missouri. After cleaning reads with the FastX toolkit, de novo contig assembly was performed with Trinity (kmer method = meryl), as aligning the RNA-seq reads to the sequenced B73 maize genome would result in loss of sequence unique to Ab10 [12]. After Trinity had assembled de novo contigs, cleaned reads from each of the four genotypes were mapped back to the Trinity contigs using Trinity's internal aligner, Bowtie [13]. The Bowtie output was converted into a usable form by Samtools idxstats [14]. A perl script identified contigs, representing expressed genes, that met the following conditions: expression in both Ab10 lines

and no expression at all in Df(L) or N10 lines. These cDNAs were potentially expressed from genes located in the distal euchromatic tip of Ab10.

Mapping Transcripts to the Distal Tip of Ab10

Given the parameters, 781 contigs were identified with expression support in both Ab10 samples, but were not expressed in Df(L) or N10 lines. Primers were designed to the 31 contigs supported by the highest read count within the Ab10 samples (i.e. most highly expressed) using Primer3 software [15]. PCR was performed on DNA homozygous for Ab10, homozygous for the deficiency line Df(L) that lacks the distal tip of Ab10, and homozygous for the sequenced inbred B73. Samples that only amplified in the Ab10 sample must be genomically located on the Ab10 distal tip. To confirm the distal location, PCR using primers for one contig (Kin618) was performed over a suite of homozygous Ab10 deficiency lines: Df(B), Df(C), Df(I), Df(F), Df(K), Df(H), and Df(M). Reaction conditions were as follows: for each reaction we used 2X Phusion Master Mix, 0.25 μM primers, 3% DMSO, and 10-30 ng DNA. Reactions were denatured at 98 °C for 5 minutes, followed by 35 cycles of 98 °C for 10 seconds, 55 °C-62 °C for 30 seconds, and 72 °C for 30 seconds, and a final extension at 72°C for 5 minutes. PCR reactions were run on 2% agarose gels.

Overgo Design and Probing of a BAC library homozygous for Ab10

A previous study used lab stock homozygous for Ab10 to make a bacterial artificial chromosome (BAC) library [16]. To obtain genomic sequence correlating with the Ab10 unique RNA transcripts, overgo probes were designed using an overgo prediction program used in Dan Peterson's lab at Mississippi State University, where the BAC library had been created [16]. For

seven of the transcripts, a single forward and reverse probe was designed. For transcript 548, two probes were designed. For transcript Kin618, six probes were designed (Table 2.3).

Library screening was performed by Lisa Kanizay using the protocol described [16]. Fifteen of the sixteen probes hybridized to the library, cumulatively targeting 94 distinct clones. While some BAC clones hybridized to only one probe, other BAC clones hybridized to as many as eleven probes (Table 2.4). Hybridization intensity varied from weak to strong.

BAC preparation, sequencing and assembly

To begin analysis, twelve BACS were picked based on the diversity of probes which hybridized to them. The twelve BACs were purified by Lisa Kanizay using the Qiagen Large-Construct Kit. BAC DNA was subsequently submitted to GGF for MiSeq PE250 Illumina sequencing. Attempted assembly of BACs was performed using Velvet and Sequencher, yielding approximately 100 contigs per BAC clone [17].

Ab10 Specific Gene Annotation

Transcripts were probed for correct reading frames using the Colorado State molecular toolkit translation tool (http://www.vivo.colostate.edu/molkit/translate/). Identification of potential functional domains and protein predictions were performed using NCBI BLASTX.

Meiotic- Specific Expression Assays

RNA was collected from leaf tissue (vegetative), root ends (mitotic), as well as anthers (meiotic), staged as described above. Tissue was flash frozen with liquid nitrogen and RNA was collected using the Qiagen RNeasy Plant Mini Kit (74904). RNA was converted to cDNA using SuperScript® III First Strand Synthesis Kit with Poly(A) specific primers. PCR was performed

on the cDNA as described above. Reactions were run on a 2% agarose gel to check presence/absence of expression.

Genomic Presence of Distal Genes in smd Mutants

Plants homozygous for the five *smd* mutations were grown and approximately 300 μg of fresh leaf tissue was harvested from 2 week old seedlings and ground in liquid nitrogen. The plants continued to grow so that we could later harvest meiocytes. Each sample was lysed with 1% CTAB, 1M Tris pH 7.5, 5 M NaCl, 0.5 M EDTA pH 8.0, and 14 M β-mercaptoethanol at 65°C for 1 hour. DNA was extracted with an equal volume Chloroform/Isoamyl alcohol (24:1). The samples were rocked gently for 10 minutes and then spun in a bench top centrifuge at 13,000 rpm for 10 minutes. The aqueous layer was extracted and treated with 10 mg/mL RNase A at room temperature for 30 minutes. DNA was precipitated with isopropanal, washed with 70% ethanol, and pellets were air dried and resuspended in 150 μL dH_2O. PCR was performed using the distal tip primers with homozygous Ab10 plants acting as the positive control. Bands were visualized on a 2% agarose gel to score presence/absence of genomic sequence in each *smd* mutant line.

Expression of Distal Genes in smd Mutants

Meiotic anthers were staged and collected from the five *smd* mutants as described above. Tissue was flash frozen with liquid nitrogen and RNA was collected using the Qiagen RNeasy Plant Mini Kit (74904). RNA was converted to cDNA using SuperScript® III First Strand Synthesis Kit with Poly(A) specific primers. PCR was performed on the cDNA as described above. Reactions were run on a 2% agarose gel to check presence/absence of expression. To

quantitatively assay down-regulation, qRT-PCR was performed for transcript 430 and 248 using SYBRGreen mastermix with an internal control of β-actin by Amy Webster. Standard dilution curves were performed for each primer to insure primer efficiency. Each reaction was performed in triplicate and fold change was calculated using ΔΔCt relative to expression in wild type Ab10 lines.

Diversity Analysis of Distal Genes

Three landraces heterozygous for the Ab10-III variant of the abnormal chromosome were grown until the 3 week seedling stage, when approximately 300 μg of fresh leaf tissue was harvested and ground in liquid nitrogen. DNA was extracted as described above and presence/absence of the distal genes was assayed by PCR. Seeds from the three lines examined were obtained from the Germplasm Resources Information Network: PI 444834 (Huila, Columbia), PI 444296 (Caqueta, Columbia), and Ames 19980 (Oaxaca, Mexico).

Results

Discovery of Meiotic Genes Unique to Ab10

In order to identify potential genes contributing to the meiotic drive phenotype of Ab10, we must look only at those expressed during meiosis, when the crucial processes resulting in drive occur. mRNA was obtained from whole-anthers containing meiocytes staged between prometaphase I and anaphase II. In total, we collected anthers from four samples: a homozygous Ab10 line, a sibling homozygous N10 line, a heterozygous Ab10 line, and a line heterozygous for Df(L). Df(L) is a variant of Ab10 that lacks the distal euchromatic tip and no longer

preferentially segregates. Adding Df(L) to the analysis controls for transcripts specific to Ab10 that do not cause drive. Using a de novo transcriptome assembly, we looked for transcripts present in the two Ab10 lines that were absent in both the N10 line and the Df(L) line.

The bioinformatic analysis yielded 781 transcripts that fit this profile. We designed primers to the 31 most highly expressed transcripts and used PCR to check their genomic presence in Df(L) as compared to untruncated Ab10. Nine of these only amplified when the entire abnormal haplotype was present, indicating their physical presence on the distal tip of Ab10 (Figure 2.2). We have now identified the first genetic sequence unique to the abnormal haplotype.

Origin and Functional Characterization of Distal Genes

Though a de novo transcriptome assembly is a valuable tool for gene discovery, the presence of multiple gene copies or alternative splicing can complicate cDNA annotation, leading to the assembly of transcripts that are an artifact of the bioinformatic analysis. When annotating genes or quantifying gene copy number, it is more reliable to base analysis on a genomic assembly. Luckily, a previous study resulted in the creation of a homozygous Ab10 bacterial artificial chromosome (BAC) library with 4X coverage of the maize Ab10 genome [16]. Though the BAC library contained the Ab10 distal tip genomic sequence, we were previously unable to probe for the corresponding clones. The discovery of nine Ab10 distal transcripts allowed us to design overgo probes specific to unique Ab10 sequence (Table 2.3). Distal-specific overgo probes hybridized to 94 BAC clones (Table 2.4). Within the group of 94 BACS, some BACs hybridized to only one probe (i.e. A2, A4) while other BACs hybridized to as many as 11 (i.e. B6, B9). Probe signal varied in intensity as either strong or weak. To begin genomic

sequence analysis of the distal transcripts, we chose 12 BAC clones that, together, hybridized to each of the 8 distal genes (the probe for gene 362 did not work) (Table 2.5). These 12 BACs were isolated and prepared for Illumina sequencing. When genotyped by PCR, the 12 BAC clones showed a different pattern of distal gene presence than when hybridized to overgo probes (Table 2.5). The "weak" signals picked up by the BAC probing experiment likely do not indicate the presence of the distal genes.

Once sequenced, the BACs could not be fully assembled, likely due to the largely repetitive content of the distal tip of Ab10. However, we did find that two of the transcripts, 430 and 248, were from the same BAC contig. When the RNA transcripts were mapped to the genomic contig, their sequence overlapped. We also observed that at least two of the distal genes, Kin618 and 405, were multicopy, the evidence of which will be presented in Chapter 3. Some of the BACs had exactly similar distal gene presence, such as C5 and C6, which both have 430, 365, and 248. We consider these two BACs to represent the same genomic area of the distal tip. This is unsurprising since our BAC library was created with a 4X coverage of the maize genome, so multiple BACs containing the same genetic content are to be expected. Furthermore, various overlaps among the BACs allow construction of a preliminary genetic map of the distal tip. For example, BAC B10 contains gene 248 as well as 548 and (B5) Kin618. The presence of these genes was determined by PCR results. Since BACs C5 and C6 share the 248 gene, their ends overlap and thus 248, 548, and Kin618 are next to each other on the distal tip. Since BAC B5 also contains (B5) Kin618 as well as gene (B5) 405, we can place (B5) 405 as the most distal gene in our map. Though the map is likely flawed, it provides a basic order of distal genes that can be further refined by future experiments (Figure 2.5).

Of the nine cDNAs identifed, only Kin618 had an open reading frame (discussed in the next chapter). The other eight cDNA contigs were riddled with stop codons. By using BlastX, we were able to find functional domains for some transcripts, which perhaps provides clues to their origins. NCBI BLAST of the non-redundant gene database indicates that the distal genes have closest homology to other maize genes ranging from approximately 82-95% percent identity (Table 2.2).

Expression Patterns of Distal Genes

Since neocentromere activity, the visual hallmark of preferential segregation, only occurs in meiosis, we reasoned that genes integral to the drive phenotype may have meiotic-specific expression patterns. To test this hypothesis, we collected RNA from mitosis (root tips), vegetative tissue (leaf) and anthers (containing meiocytes) of a plant heterozygous for Ab10. Of the 9 transcripts residing on the distal tip of the Abnormal chromosome, five are meiotic-specific. Three are expressed in all three tissue types and results from one, 615, were inconclusive (Table 2.7).

smd3 and smd8 are Terminal Deletions

To determine whether any of the distal genes play an important role in the drive mechanism, their genomic presence was assayed across mutants deficient in meiotic drive. The *smd* mutants were created in a forward mutation screen from a line with active *Mu* DNA transposon. The lines isolated are Ab10 chromosomes that no longer preferentially segregate. In *Mu* mutation screens, the *Mu* transposon theoretically jumps into the gene and disrupts transcription, thus causing an altered phenotype. Five independent mutation events, *smd1*, *smd3*,

smd8, *smd12*, and *smd13* all have reduced or absent drive phenotypes, but the cause of the mutations has not been identified. Interestingly, both *smd3* and *smd8* lack genomic copies of several of the distal genes, indicating that the mutation events are distal deletions (Figure 2.3, Table 2.8). *smd3* lacks six of the nine distal genes and is a far more extensive deletion than *smd8*, which appears to only lack 548 and one of the copies of Kin618 and 405 The *smd* deletions provide important data confirming the order of the genes on the distal tip as predicted by BAC clone content and overlap (Figure 2.5).

Transcripts 430 and 248 are Absent or Downregulated in Every smd Mutant Line

Though the *smd3* and *smd8* mutations are a product of terminal deletions, the mutant lesions in *smd1*, *smd12*, and *smd13* remain unclear. Kin618 is downregulated in *smd12*, a fact that will be addressed more fully in Chapter 3. Interestingly, transcripts 430 and 248 showed downregulated expression in *smd1*, *smd12*, and *smd13*, while all the other distal transcripts show no such pattern (Figure 2.4, 2.5). As both 430 and 248 are absent in *smd3* and *smd8*, these two transcripts are thus physically absent or downregulated in every Ab10 mutant line.

Seven Distal Tip Genes are Conserved Across Ab10 Types and Geographically Disparate Landraces

The presence of Abnormal chromosome 10 has been assayed across both landrace and teosinte lines and found in all assayed populations at an overall frequency of approximately 15% [16]. There are also three abnormal chromosome 10 types varying in knob content, Ab10-I, Ab10-II, and Ab10-III, that preferentially segregate [16]. Any distal transcripts involved in the meiotic drive phenotype should be conserved across geographic, subspecies, and abnormal type varieties.

The great genetic diversity in the species *Zea mays* means that conservation of genes likely points to functional significance. To perform this investigation, we scored the presence or absence of the nine distal genes by PCR in three geographically distinct landraces containing Ab10-III isolated from Huila, Columbia, Caqueta, Columbia, and Oaxaca, Mexico. Seven of the transcripts were conserved across these three lines. Two of the distal transcripts, 365 and 348, were not present in the Oaxaca line, indicating they likely are not involved in the drive phenotype (Table 2.8). These data indicate that all of Ab10 haplotypes assayed have a common origin and may suggest that some of the transcripts have functional roles.

First Ab10 Distal Genetic Map Shows Gene Duplication and Expression Clustering

To date, genetic mapping of the distal tip has proven impossible because of both the lack of markers and Ab10's inability to recombine with any other chromosome in the genome. By combining the information about the PCR-verified presence of the distal genes on BACs and the deletion mutations *smd3* and *smd8*, we are able to construct the first genetic map of the distal region of Abnormal Chromosome 10 (Figure 2.5). Two of these transcripts, 405 and Kin618, are duplicated and a cluster of meiotic-specific transcripts lies between the duplication. Though parts of the nine distal transcripts retain homology to other parts of the maize genome, their closest homologs are scattered across the nine chromosomes and do not come from any one particular region, indicating that the origin of Ab10 was not a simple duplication of part of the maize genome (Table 2.2).

Discussion

The Rhoades model of meiotic drive is an impressive theory that suffers from a lack of molecular support. The greatest hindrance towards developing molecular evidence of drive was our lack of knowledge of the sequence unique and specific to the abnormal haplotype. Neocentromere activity and the preferential segregation of knobs only occurs when trans-acting factors located on Ab10 are present in the plant [11,18,19]. A series of Ab10 deficiency lines placed the meiotic drive factors on the distal tip of the large chromosome [6,10]. Since neocentromere activity occurs during meiosis, we used high-throughput sequencing to target the meiotic transcriptome of a plant with Ab10 to look for unique mRNA transcripts. By comparing transcripts between plants with Ab10, the distal deficiency Df(L), and N10, we identified the first DNA sequence unique to Ab10, providing a valuable gateway toward understanding the abnormal haplotype.

The previous creation of a BAC library homozygous for Ab10 allowed us to design overgo probes based on Ab10 transcripts and pull out BACs containing Ab10 genomic DNA. Assembly and annotation of these BAC sequences provided the first glimpse of the unique abnormal haplotype. Though past studies have hypothesized Ab10 originated from a wide cross with a separate grass species, the genes of the distal tip retain closest homology to other regions of the maize genome. Rather than being a simple duplication of one chromosomal segment, the distal tip contains an amalgamation of other parts of the genome that likely originated from the last whole genome duplication event. Indeed, the paleopolyploidy event in maize resulted in much rearrangement across the genome and would have provided an ideal opportunity for the abnormal chromosome to arise [20]. The series of nested inversions just proximal to the distal tip

that prevent N10 and Ab10 from recombining have retained the integrity of the extrachromosomal segment and maintained the linkage between the target knobs of Ab10 and the driving genes located in that distal region, causing the distal region to evolve separately from the rest of the genome. The distal tip itself is full of duplications and transposable elements, which makes sense in a region that rarely recombines. Further study may reveal how a selfish system exists in a region so susceptible to genomic disruption.

This study identified nine meiotic drive gene candidates residing on the distal tip of Ab10. Interestingly, among the nine transcripts we identified, only one coded for a fully functional protein. A further diversity analysis revealed that two of the nine genes are not required for function. However, the six other transcripts remain strong candidates for playing a role in drive. Though Kin618, the C-terminal kinesin, is likely involved in moving Knob 180 neocentromeres and will be addressed more fully in the next chapter, meiotic drive is a complicated process where multiple events must occur. In order for Ab10 to preferentially segregate, there must be a recombination event between the knobbed haplotype and the centromere. Furthermore, the knobs must maintain polar localization throughout meiosis I and meiosis II to ensure inclusion in the most basal cell[18]. The distal transcripts identified in this study may possess an RNA-mediated function that increases recombination near heterochromatic knobs or provides a nuclear envelope localizing signal to maintain polar orientation. Also, Kin618 may not function independently to move neocentromeres, but rather work in tandem with another distal transcript. Kin618 does not have meiotic specificity (Chapter 3) even though neocentromeres are only observed during meiosis (unpublished data). Some of the distal transcripts may provide either knob DNA-binding or meiotic specificity to Kin618. Transcripts 430 and 248 particularly show promise as integral contributers to the drive phenotype as they are

downregulated or absent in all *smd* mutants, even those in which no other known aberration has been observed, such as *smd1* and *smd13*. Together, the seven transcripts and the Ab10 genomic BAC sequence are potent molecular tools that target the 70-year-old mystery of Ab10 meiotic drive.

Acknowledgements

We would like to thank Alex Harkess for bioinformatics help and advice with the Trinity *de novo* transcriptome assembly. Lisa Kanizay provided extensive intellectual guidance and practical help including sequencing RNA from meiocytes in a previous trial, probing the BAC library for distal tip genes, and preparing targeted BACs for the first round of Illumina sequencing. Amy Webster conducted qRT-PCR experiments for the expression of the distal tip genes across the *smd* mutants.

References

1. BURT, A., Trivers, R. & Burt, A. *Genes in Conflict: The Biology of Selfish Genetic Elements*. (Harvard University Press: 2009).at <http://books.google.com/books?id=8e8xLbmtuMcC>
2. Rhoades, M. M. Preferential Segregation in Maize. *Genetics* **27**, 395–407 (1942).
3. Ananiev, E. V, Phillips, R. L. & Rines, H. W. A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10785–90 (1998).
4. Peacock, W. J., Dennis, E. S., Rhoades, M. M. & Pryor, A. J. Highly repeated DNA sequence limited to knob heterochromatin in maize. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 4490–4 (1981).
5. Kato, Y. Cytological studies of maize (Zea mays L.) and teosinte (Zea mexicana Shrader Kuntze) in relation to their orIgin and evolution. *Mass Agric Exp Sta Bull* **635**, 1–185 (1976).
6. Hiatt, E. N. & Dawe, R. K. Four loci on abnormal chromosome 10 contribute to meiotic drive in maize. *Genetics* **164**, 699–709 (2003).
7. Hiatt, E. N., Kentner, E. K. & Dawe, R. K. Independently regulated neocentromere activity of two classes of tandem repeat arrays. *The Plant cell* **14**, 407–20 (2002).
8. Mroczek, R. J., Melo, J. R., Luce, A. C., Hiatt, E. N. & Dawe, R. K. The maize Ab10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. *Genetics* **174**, 145–54 (2006).
9. Lyttle, T. W. Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends in genetics: TIG* **9**, 205–10 (1993).
10. Kanizay, L. B. The variants of maize chromosome 10 and their roles in meiotic drive. 111 (2011).
11. Dawe, R. K. & Cande, W. Z. Induction of centromeric activity in maize by suppressor of meiotic drive 1. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8512–7 (1996).
12. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**, 644–52 (2011).
13. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357–9 (2012).
14. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078–9 (2009).
15. Rozen, S. & Skaletsky, H. Primer3. (1998).
16. Kanizay, L. B. *et al.* Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in Zea mays. *Heredity* **110**, 570–7 (2013).
17. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* **18**, 821–9 (2008).
18. Rhoades, M. Preferential segregation in maize. *Heterosis* 66–80 (1952).
19. Yu, H. G., Hiatt, E. N., Chan, A., Sweeney, M. & Dawe, R. K. Neocentromere-mediated chromosome movement in maize. *The Journal of cell biology* **139**, 831–40 (1997).
20. Wei, F. *et al.* Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLoS genetics* **3**, e123 (200

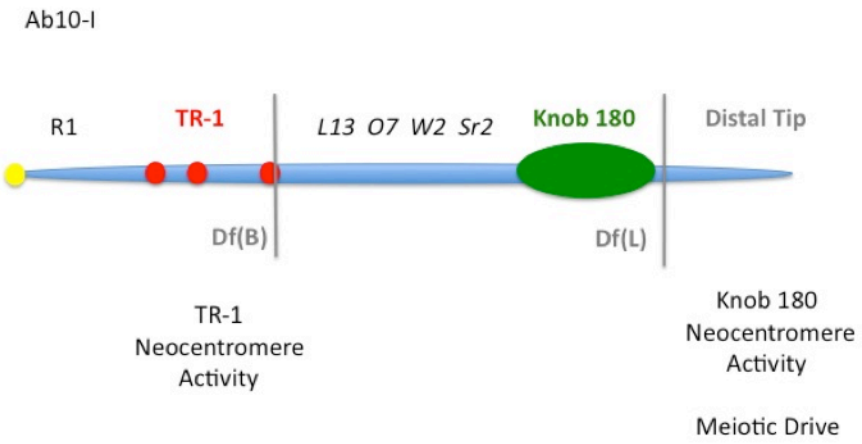

Figure 2.1: The long arm of Ab10-I with its distinctive pattern of knobs. Yellow marks the centromere, red marks the three TR-1 knobs and green represents Knob 180. Df(B) and Df(L) are Ab10 deficiency lines that lack distal portions of the abnormal chromosome. The trans-acting factors causing meiotic drive and Knob 180 neocentromere activity map distal to the Df(L) breakpoint.

Table 2.1: Summary of high-throughput sequencing data presented in Chapter 2.

Files	Genotype	Tissue	Library Creation	Sequencing
AB10_s_3_1_GCCAAT_sequence-paired-trimmed-paired.fa	Ab10/Ab10	anther RNA		454 (Lisa Kanizay)
RNJ_s_5_1_CAGATC_sequence-paired-trimmed-paired.fa	N10/N10	anther RNA		454 (Lisa Kanizay)
GGF_08_SOL_0001_GCCAAT.fq	Df(L)/N10	anther RNA	TruSeq RNA	Illumina HiSeq SE100
GGF_08_SOL_0001_TGACCA.fq	Ab10/N10	anther RNA	TruSeq RNA	Illumina HiSeq SE100
A9	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
B2	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
B5	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
B10	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
C4	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
C5	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
C6	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
E9	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
F6	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
G11	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
H4	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
H5	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250

>kin618(B5)

CTGCCGGGACCGGGAGGCAGCCACTCGCGGCCGCGCCACCGCCGCCGGAGGTCCCC

CTTTCCAGTTCCATCCGAATCGAGAGGCGGAGGCGCCCGGATCCACCGCTCACCCCC

GCTCCTCAAACCTCCATGGATCCACCCGCCATGCCCGCGCGACGGAGGCGCCCGGA

GGTTCTGCACGACAAGGAGAACCCGGCGGACGCGCAGCCCAGAGAGCGCCAGCGC

ACCGCTGCCGGGACCGGGAGGCAGCCACTCGGGGCCGCGCCGACGCCGCCTGAGGC

CCCCCTTTCCAGTTCCATCCGAATCGAGAGGCCGAGGCGCCCGGATCCACCGCTCAC

CCCCGCTCCTCAAACCTCCATGGATCCACCGCTCACCTCCGCTCCTCAAACCTCCAT

GGATTCACCCGCCATGCCCGCGCGACGGAGGCGCCCGGAGGTTCTACACGACAAGG

AGAACCCAGCGGACGCGCAGCCCGGAGAGCGCCAGCGCACCGCTGCCGGAACCGG

GAGGCAGCCACTTGCGGGCGCGCTGCCGCCGCCGCCGCCGGAGTCTTTGCAAAATG

AAGCTGATTTATTGAAGGAGGTTGAAAGTCTTAAAATCGAGCTTCAACATGTAAGGC

AGGACCGTGACATGTTACTGTCTTATATGGGAACATATAAAGAAATGACTGAAAAA

TCTGTCATTGAATTAGAAAACGCCATGAAAAAAGCCTCTACCCTTCAGGAAACATGT

TCGTCCCAGAGAAATATTATAAGAACATTGGAAAGTAAGCTTGCAACTGTGGAGAA

GTTAAAGATGTCTAATATGATGTCCTCAGAAACCATGACTGAGTATGAGGACATGA

AGAAATTGCTAGAGAGTGTCCAGTTACGCCTTAAAGAGGCTGAGCAAACAATTTTG

AATGGAGAGATTTTACGCAAAAAGTTACATAACACAATTCTTGAGTTGAAGGGAAA

CATCAGAGTATTTTGTAGAGTAAGGCCTTTGTTGCCAAATGAATCTGGAGCGGTCTC

ATATCCAAGGAATATACAAGACATAGGCCGAGGCATTGAGTTGTTACATAATGCAC

AAGGATATTCATTTACTTTTGATAAAGTATTTGATCATTCCGCAACGCAGGAACACG

TATTCACAGAAATATCTCAACTCGTGCAAAGCGCTCTAGATGGCTATAAGGTCTGTA

TATTTGCCTATGGGCAAACTGGTTCTGGTAAAACATACACAATGATGGGCAACCCTG

AGTTAGAAGAACAAAAAGGAATGATACCACGATCTCTGGAGCAAATTTTTCAAGCT
ACCCAGGCACTTAACTCACAGGGATGGATATATAAGATACAGGCCTCAATGTTAGA
AATTTATCTGGAAACCATACGTGACTTGCTTGCGACGAAACACATTTCTCAAGATGG
TGGTTCTTCAAAGCACAATATCAAGCATGACGCACAGGGTAGTATTGTGTCGGGCCT
TACGGTCATTGACGTAACAAGTATAAGCGAAGTTTCTTCTCTCCTTAGGCGGGCTGC
TCAGAGCAGATCAGTTGGGAGAACACAAATGAATGAGGAATCATCCAGAAGTCACT
GTGTTTTAACTCTACGGATTTTTGGTGTAAACGAGAGAATGGACCAACAAGTGCAAG
GAGTGCTCAATCTTATAGATCTAGCAGGCAGTGAGCGCCTGAACAAAAGTGGTGCC
ACTGGTGACCAGTTGAAAGAGGCTGTGGCCATTAATAAAAGCTTATCATGTTTAAGT
GATGTCATCTTCTCAATCGCAAAAAAAGATGAACACGTTCCGTTTCGGAACTCGAAG
CTGACCTATCTGCTCCAGCCATGTCTCGGAGGGGACTCCAAGACGCTGATGTTTGTG
AACCTATCCCCTGAAGTGTCTTCGACTGGGGAGTCTCTCTGCTCATTGCGGTTTGCTG
CGAGGGTGAACTCGTGCGAGATCGGGATCCCTCGGCGCCAAACCAGAATGTCGTCT
GACCGACCAAGTTGAAATAGGACATCAAGTACTCACAATGCTCGAGGCAAGAGGGT
GGAATTCCATGAAACAGAGCAAATGAACGAGAGAGCGACTGCTAAAGGACATGAA
GTTGAAGCTGGATGAAGGAAAAAACTACCTGCAACCGATGGCAATAATGCATTGCA
GGGTTGAGGCTCGTCAGGAA

>548

AGTTCCATTTCTGTTCAACTGATGTTCCTGAAACCAAGAAGTCTCTGCATGTCACACT
TACCATGAATGATATCCAACTTCAACACTGTGCACTGCATACTTTTTTCGGACAATC
ACGAGCTTTTTTTGTTTCAATGAAAGCCACATTTCATATACATAGAAAGTGGATTCTT
CTGTATATCCTTAATTTTCTCTAGCCACTGCTATGCAAGAACTCATACACAACATTGA
CAGTTATCTTCTTAAACCGAGTAGGCGAATATGATGCACACCAACAACGGTAAGAG

TTTGGAGCATCTTCCTCATTGGATTTTTCTCTGCTTTTTTTGATCTATTCAGTTCATAT

GTTAATGAACTGGCTAAGTTTATCAGTGCTTTGTTATGTTATTTACTTTGCATTTAGC

TTATAATAGAAATTACTCCTTCGTTTCAGATTATAAGATGGTTTGGATTTTCTAATAA

ATTATTCTTATTACTTGGTACATAACTAAGAGTAGAACTGATAAATTAGACAATGTG

TTTATTTGTGATTGTATCCACTATTTATATTATGTTTAAATAATAACAAGGTTTATGA

ATTGATATTGCAATATAGCTATTAGTTACCTAGATTGTCCTTGCGAGCATTTTCGTGT

GCCAACAAATCGTAATTCAGGCTACTGAATATAGGTGCCCATATGCCCCATCTTGAT

TTTCTGTCATTATTTATTGACCCTCTTTTATCTTTGATCTACCATACACTACAGATTTA

AGATTTACATATTTTACCACGCTGTGCCTTGAAGCCCTTGATCACGCGCCGCCGGCC

ACGCTGGTACCCCGACAGTCTCCGCCCATGACCAATTAATCACGAAGCCATGGAGG

CGGATGTCGCTGACTCACCACGACGCCTTACTGAGCTACTGATGCCGGCACACACCG

CTCCTCATGGTTCCTAATTGATCTATTCTATTATGCCAGTAACGAGTGTCCGAATGGT

GGTCCTCCCTCCCGTGGAGGTTATGGGCGCGACAGAGGGGGCCGTGGGGTTGGTGG

AGGCCAGCGAGCTAGTGCCCGGGAAACTGAGGAAGACCAAAAGCATAATGAGAAA

GCCCCTCTTCCTAAGAGGCAAGATGCATTGGAGTATAGTCCACAGGTGGAGCAGAA

AGCCCCTCTTCCTGAGAAGCAGGGTGCAGTGGACACATATATTAATGAAGAGGATA

AGGCAGATTCTCCAAAGGTCAGTGAACAAATTAGATCGGTTTATAATTCAATGAAGC

AGCTCAATGATATGATATCAAACATAGAAAAACAAGAAATAGATTTCCAATCCAAC

GATGAGGCTTCCAACACAGGTGCAAAGGAATCTATGACTAAAAATTGTTTGATGTTT

TTAAATACCGCCACATTTGAGGAGTTAAAAAGTCTAAAGGGTATTGGAGATAAAAG

AGCGAAGTTTATACTTGAACTTCGAGAGGAGTCACGGGAACCCTTCACAGAGGTTGT

TGATCTTAAAAATTTCTTGGGGATGAATAAAAGAGAAATCACCAAAATGGTGTCAG

GGATGGTATGGACAAGCTGAACGAGGAAGGTGCTGTATCAGACAGAACCATTTAGG

TTTGATGGTTTGGATGAAGTGGCGGATGCCTCTGAACTGGCGGGTGTCTGTACGCCC
GATATCAGCTGCAAGGCGAGTGTGAGCAATGTCGAGTGGGTAAAGAAGTAGCTAAG
AGCAATAGAAGTGAACTTATTGTCTGATGAGGTTCCAGCATCTTTCAGTATGCTCCT
GTACAGATCCTGCAAGAATAATGGAGACAGGTTACAGACAATGATCCGAGTAGACG
AGTTGTGACTTGTGGGAGGAGGATGCCCTTTCAACATCACAAATACTCATCAAGTTT
TCTGTCTCATGTACTCTTGATAAATCAAGAAAATTGAACATAGGTACTCATGCACAC
GGAAACTATAATATGCTAGCTGGAGGTCGATGCCAACGTCATCTTAAAAACATGAA
AGTCATCTTAAAAACATGAAAGGTAGTGCAATCTACAGTGCAAGTTAAAGTTGTTCT
CCAAGGCATTGACCGGATATGGTGCATCATTTGGCGTTCTAACCATAAACATACATT
AAATTCTAACTGTATGCTTTTCAAACCTTGTGCAGGTAAAGGATATCAGATACATTA
ATGTATACCATTGGTTATAGACTTATAGTGTTGGTAAAACACAGAAAGTATG

>405

TTGAGTCAAAGATTTAACATTTGCATTGTGATTTACTAGGAATTTTCTAACCATGATG
TCTTTTGTATAGCCTTTTTGATGTAAACAAATTTACTTTGGCAAATTCCAGCATCTAT
TGGGGCAATTCCATTACTTACACTTACAGTACTGTAAGTGATCATAAAACTTCTATG
GTTCAGGAAGCTGCTTATATGTTGTGCTCTCTGAATGAATACAGTCTTCTAAACGAC
AACAGTTTCATGGGACCAATTCCTACTCTCAACAGCGTGACTAAGCTCAAAGTTTTG
TGAGTCGTACAATTACAACATTTAACTTTATTAGGACAACATCTGTGTTATGTCAACT
GTAAATTTTTTTTTTGTGGAAATAACAAAATGCTTTCAAATAACAAACTAACTGGAC
CAATTCCTAATTTGACCAGGATGGGCTTGCTGGAAATTGTGTTAGAATTTCTTTCAA
GCTCGCCAAATTGTGTGTCAGTCGCAGCGCACAGGCATAGTCCCAGCGGTTCTTTAT
TTATTTGTTGCAAATGTCTTCTCTCTATGTGCTCGCACCACAAAATTCCAAA

>430

GAAAATCTGAGACTAGTTAGGAAGCTGGATCCTAGCCCATGCTGTGGTAAACAAGC

CATCTTCTATTCACTCATATATGGACATGTGGGTTACACATATTTTCCACAGTTCTCA

TCCTGGCAGGAAAAGGCAAAAACTAATGACCTCAAAATTGGTTTATCGATTGAGTC

AACATACGAACATAGCCGTACGGTAGGTGAGGGAGAGATTTTCA

>365

AAAATGTCACTTGATTGAAATAATGTTTAAACGTTGTACAGTAAGGTTCGTAGACGT

ATTAGATGTCAAACTTTGGTAAGAATAAAAAAGAGAAAAAACCATGTGCATTATTG

CAGGTACCATAGCACATAGTGAATTTGTGAGAAAGTAGCCGTAGCTTGTAGTTGTTG

GTCACACACTAAGACCAAGGAGCTAAGCCGATTCACTCGGACTGTAATCTCAAACG

CTAACTCCTTTCCTACAATGACAAAGAGATAAACATGTCATGTGATCTTAAAGGAAG

TGGTGACAATGGCAATGTTCCATTCCTTTTTTTTTGCATTGCATTTAATGTTGGCATA

TATAATCGCACAATGTAAGTGGATGTAGATGCTGTAAATTTGGAGAATTTGGGAAG

ATAAGTACCTGGCATATACA

>248

AGAATGGAATTCAATTCCAATGACCCAAACGGGGCGTAACTGTATCTGCCAGGTCTT

GGAAATTAATCTCATGCTCAGGGAGGGGTGCTATACTGCTATCTGAACTATCGCCCA

ATCATCCCGGTGTTGAACATGTTCAGCCACCAAACCTTGCGGGAGGAATGGTGTAGG

ATCAACACGATGGTGCGCCATTGCTGAGTTCTGCACTGCTTCTTGAACTGGGGTTTG

GAGAGGATGAGTTGGGGATGCTGGTTGCTTTCTTCAATGAACCAGGAGACGAGTTA

CGCGTCGAATGGAGTTCGGTGAAGAGTCACCCCTAACTTGCTGGTAACTTGAACCAT

GAGGAAGGGCCGATTGCCATTTAGCATTTACAAATTTGTTATGGAAGGATGCTGAAA

ATCTGTGACTAGTTAGGAAGCTG

>615

ATATAGTATATATATATACTATATATATATATATAGGTCCTGGTACATATAGGATGA

CTAGGGTCGACTAGGCTCAATTAATTGGGCTTATAAATGACCTTGAATAGAGTGATA

CTCCACCTGAGAAGATAAAAGTGGATGTCTTCAGGGCAGTGGTGATGGCCGTTTAAT

GCTGCTATGGAGATGGTGCAATGAGGGGACTCAGCTTTGCTGCTCCTTTTCCTTTCTC

ATTTTGGTTGGTGAGCTTTTAACTCAGAGCTGTAAGAGGATGTTTCTTGTCTTTTATG

AATATTAAAGGCTAAGGAGACCCTGCCCTCGAACTCTAAAAAAGGGCTTATCGATG

ACTAATCACATTATTGGTGTCATGGCAACTAGGATCAATTAGATGATTTGAAAAAAT

TTCACAGAGTTATACCTTCTCCATTCCCTTGACCA

>348

GTTTAAACCATCTGTTCAAGACAAACGTGTATTCTGAAAAACAAAAGACGACTATTG

CATGTCTTCTCCTTCATGCATGACGAGGTAACTCGAAATACTTCCAGGTTTTCAACA

AGCATATGAACACAACACTGTCATACTACATGACTATCAAAAACAAAGTGCACCTTC

TCCATCCTATTTAGACAACTCTCGGTTCAGCCTTCTTTTGTGCAACAGATTCCTGCAG

ACCAGTCATCCTCTCCAAGAGTCCATACAGTGTTTGCTCAAGCCGGGCATTCAAAAA

TCTGTACTGTAAGTAGCGATTGTGTAGAATCCGCAACCTATGCGCATCCTCCTGCTTC

TCCGCATCCTTGTTCCGTGCAGGCACAGGCACAGGGAACAATGATTTGCGCGTTCGT

GTCGATGGCCGCATACAACCCCTAGGGTGGTGTTCGTGCGCCGTCGTGGTGG

>362

GGGAAAAAAAGGCTGCTTGATTAAGGTACATTTCACTACGCAATCCTATCCCTAACA

GTGCCTTCCAATTTAACAATTTCTCTTCAGTGGTGGTACTATTACGTATTTATTATTT

GCATAACTATTTCTGTATGTACTTCTTCTTTGCATAACTATCTATTGCTTGGTACTTTT

ACTCAAAACATTTCACAAGTTTGGTACATTACCTCTAAACACAACGTTCATTTAATA

TCTATATGTATTACTTGATACATTTATTGAAAAACATTTCACAGGTTTGCTACATTAC

CTCTAAACACAACATTCATTTAGTATTTATAGTACCTTAGTAGTTTCGCACAGAAAC
GTCCGTGCATTGACTGCTTACTGTTTTGGTATTTTCAAATTGACAGTAAATCCTAATA
TTTGGTTTCGAAACATGAGCTTTGTCAACTTATACATATCTGCACAGGGGAAAAAAG
CTAATTGACCAAGATACAACTCATTACACATTCATTTTCATAACAGTGTGCTCCATTT
TTACTATTTTTTCCTAATAGCGCCCACCTGGTTATTCGCTAAAACATGTCTTATGTAT
CACATATGATTGCTTAATAGTTGGCTTATTTTTTTTTCTCAAACGCGCAGG

>613

GCATTTGCTAGGGTCAGGGCTATTTCTAGTGCCTGATGCTGTATTGTTTTTCCTTTTG
TTTTTTTGCCTGTTATGGCATTGTACTTTGGTCCACTTCGGACATTTCTCTCTTTCAAT
ATAATGATGTGCAATTCTCCTGCGCGTTCGAGGAAAAAAGACCATTAGCAACTTTTG
ACCCAGTCTATCAGTTTTGGGGTTCAATCCAACTCATGGGTGAAGTTTTGGTGAGCC
AGTAACCTCTCTGTTTTGGTATTTAACTTCATATTATTGTCTGAAATTGTGATGTATT
GGTATTTTCCTATGTTGTTTAAAACTATGTTTAGTAAAAATATCACCCATGAGCGTTC
CAACGTTT

Figure 2.2: Fasta sequences of Ab10 distal tip expressed transcripts predicted by Trinity De Novo assembly. The Kin618(B5) sequence has been further refined, as explained in Chapter 3.

Table 2.2: The region and percent identity of closest homology between the Ab10 distal expressed transcript and the corresponding region of the sequenced B73 genome as predicted by MaizeGDB BLASTn. The BLASTX results represent the distal transcripts analyzed in the non-redundant (nr) nucleotide database of NCBI.

Distal Gene	Location of B73 Homology	Percent Identity	BlastX Result
Kin618	chr7; 1144960-1144796	95.15	C-terminal kinesin
548	chr10; 16321061-16320645; 19345602-19345284	94.48	kinesin-like protein KIF-22 like
405	chr8; 112899007-112899500	82.39	receptor protein-kinase like
430	chr4; 4734449-4734307	83.92	WRKY domain
365	chr5; 211354016-211354226	82.46	unknown
248	chr3; 88508899-88508695	87.8	hypothetical protein: Zea mays
615	chr2; 6345080-6345319	95.83	unknown
348	chr1; 134596482-134596251	89.66	DUF566-domain containing protein
362	chr3; 130001167-130001772	96.2	unknown

Table 2.3: Forward (F) and Reverse (R) probes used for probing the Ab10 homozygous BAC library. Probes labeled "614" target Kin618.

Probe	Sequence
615-2.F	GTCCTGGTACATATAGGATGAC
615-2.R	CCTAGTCGACCCTAGTCATCCT
248.F	CGGTGAAGAGTCACCCCTAACT
248.R	TCAAGTTACCAGCAAGTTAGGG
348.F	CAGACCAGTCATCCTCTCCAAG
348.R	ACACTGTATGGACTCTTGGAGA
430.F	GAGTCAACATACGAACATAGCC
430.R	CTCACCTACCGTACGGCTATGT
362.F	AGCTTCAGGCATCCCAACGTTG
362.R	CCAGTAGATTCCAACAACGTTG
405.F	GGGACCAATTCCTACTCTCAAC
405.R	AGCTTAGTCACGCTGTTGAGAG
365.F	CACACACTAAGACCAAGGAGCT
365.R	GAGTGAATCGGCTTAGCTCCTT
614-1.F	TCAGCCTCTTTAAGGCGTAACT
614-1.R	GCTAGAGAGTGTCCAGTTACGC
614-2.F	TCTGGGACGAACATGTTTCCTG
614-2.R	AAGCCTCTACCCTTCAGGAAAC
614-3.F	GACCGCTCCAGATTCATTTGGC
614-3.R	GTAAGGCCTTTGTTGCCAAATG
614-4.F	GTAGAGGGCCTATGCTACACCT
614-4.R	AGCCACATACATAGAGGTGTAG
614-5.F	TTTATTTGTTGCTCTAGCTGTC
614-5.R	GGTTAGCGGCGGGAGACAGCTA
614-6.F	GCTCCTGCGTTGGTGACTTGGT
614-6.R	AGTCATCATTGAACACCAAGTC
548-1.F	CAACGATGAGGCTTCCAACACA
548-1.R	GATTCCTTTGCACCTGTGTTGG
548-3.F	TCAGGCTACTGAATATAGGTGC
548-3.R	GATGGGGCATATGGGCACCTAT

Table 2.4: Probing results of the Ab10 homozygous BAC library showing the 94 BAC clones and their probe-specific pattern. The 94 clones represent a sublibrary from the full BAC library. Black, bold numbers represent strong probe signal and grey numbers represent faint signal.

	1	2	3	4	5	6	7	8	9	10	11	12
A	548-1, 3	348	548-1, 3	615-2	615-2	548-1, 3	348	615-2	614-1, 5 614-2, 3, 4, 6 548-1, 3 405 430 615-2	548-1, 3	548-1, 3	348
B	614-1-6 548-1, 3	614-1 - 6 548-1, 3	615-2	248 348 365 405	614-1 - 6 548-1, 3 430	614-1 614-2 - 6 548-1, 3 405 430 615-2	365 248 348 405 430	548-1, 3	614-1, 5 614-2, 3, 4, 6 548-1, 3 405 430 615-2	614-2, 3, 4, 6 548-1, 3	548-1, 3	
C	548-1, 3	548-1, 3	614-1 614-2 - 6 548-1, 3 405 430 615-2	614-1 614-2-6 548-1, 3 405 615-2	365 248 348 405	365 248 348 405 430	615-2	615-2	348	614-1 614-2 - 6 548-1, 3 405 430 615-2	348	615-2
D	614-1-6 548-1, 3	614-1, 5 614-2, 3, 4, 6 548-1, 3 405 430 615-2	615-2	615-2	548-1, 3	614-1 614-2 - 6 548-1, 3 405 430 615-2		548-1				548-1, 3
E	348	348	348	248 348 365	548-1, 3	614-1 - 6 548-1, 3	614-1 614-2 - 6 548-1, 3	615-2	614-1, 3, 5 614-2, 4, 6 548-1, 3	548-1, 3	615-2	615-2

							405 430 615-2		405 430 615-2			
F	548-1, 3	614-1 - 6 548-1, 3	614-1 -6 548-1, 3	614-1- 6 548-1, 3	615-2	614-1 614-2 - 6 548-1, 3 405 430	348	348	548-1, 3	348	615-2	348
G	615-2	614-1 614-2 - 6 548-1, 3 405 430 615-2	614-1 614-2 - 6 548-1, 3 405 430 615-2	548-1, 3	548-1, 3	348	348	615-2	348	548-1, 3	614-1 614-3 405 615-2	548-1, 3
H	615-2	548-1, 3	615-2	615-2	348	615-2	614-3 548-1	615-2	548-1, 3	615-2	Empty	Empty

Table 2.5: Gene content of the 12 sequenced BACs predicted by PCR using distal-specific primers.

	A9	B2	B5	B10	C4	C5	C6	E9	F6	G11	H4	H5
Kin618	x	x	x	x	x			x	x	x		
548			x	x								
405	x		x		x			x	x	x		
430						x	x					
365						x	x					
248				x		x	x					
615											x	
348												x

Table 2.6: Primers used for assaying both genomic and transcriptomic presence of Ab10 distal genes. Beta-tubulin primers were used as the housekeeping gene control for qRT-PCR analysis.

Distal Gene	Forward Primer	Reverse Primer
Kin618	GCCTCGGCCTATGTCTTGTA	GCCTTAAAGAGGCTGAGCAA
548	GGCAGATTCTCCAAAGGTCA	CTCCTCAAATGTGGCGGTAT
405	CAACAGTTTCATGGGACCAA	TTGGCGAGCTTGAAAGAAAT
430	CCCATGCTGTGGTAAACAAG	AAATCTCTCCCTCACCTACCG
365	GCCGTAGCTTGTAGTTGTTGG	GAACATTGCCATTGTCACCA
248	CAGGGAGGGGTGCTATACTG	CCCAACTCATCCTCTCCAAA
615	GATGTCTTCAGGGCAGTGGT	GCAGGGTCTCCTTAGCCTTT
348	TGCAACAGATTCCTGCAGAC	CGCAAATCATTGTTCCCTGT
362	TCCGTGCATTGACTGCTTAC	TAACCAGGTGGGCGCTATTA
beta-tubulin	GATCCCCAACAACGTCAAGT	CTGGTACTCGGACACGAGGT

Table 2.7: Tissue specific expression profiles of 8 of the 9 Ab10 distal tip genes. Testing for gene 615 was inconclusive. PCR was performed on cDNA and assayed for presence/absence. Five of the nine genes are meiotic specific and not expressed in either vegetative or mitotic tissue.

Expression in Ab10/Ab10	mitosis (roots)	vegetative (leaf)	meiosis (anthers)
Kin618	x	x	x
548	x	x	x
405	x	x	x
430			x
365			x
248			x
348			x
362			x

Table 2.8: Genomic presence of the distal tip genes in homozygous *smd3*, homozygous *smd8*, and three geographically diverse landrace *Zea mays* lines containing Ab10. *Smd8* is a small distal deletion that lacks 548 and the B5 copy of Kin618 (see Figure 2.6 and Chapter 3). *Smd3* is a much larger distal deletion that has lost 6 distal genes. Genes 365 and 348 are both not found in Ames 19980, which contains a variant of Ab10-III.

	smd3/smd3	***smd8/smd8***	**PI 444834 Huila, Columbia**	**PI 444296 Caqueta, Columbia**	**Ames 19980 Oaxaca, Mexico**
Kin618	x	x	x	x	x
548			x	x	x
405	x	x	x	x	x
430		x	x	x	x
365		x	x	x	
248		x	x	x	x
615	x	x	x	x	x
348		x	x	x	
362		x	x	x	x

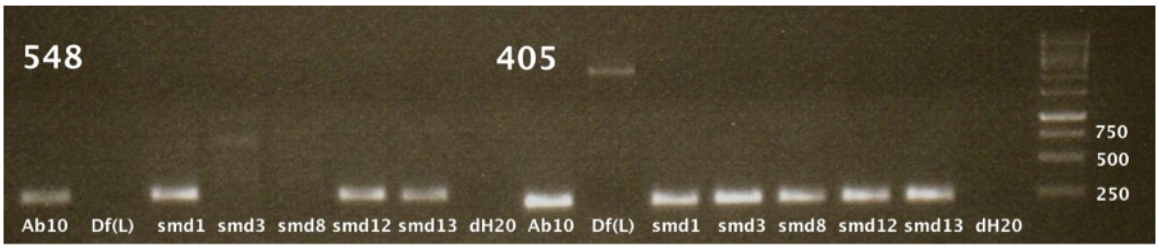

Figure 2.3: Genomic presence of distal genes 548 and 405, showing that *smd3* and *smd8* are deletion mutants lacking certain distal transcripts. Amplicon 548 is approximately 350 base pairs and amplicon 405 is approximately 300 base pairs. Size standards provided by GeneRuler 1 kb DNA ladder.

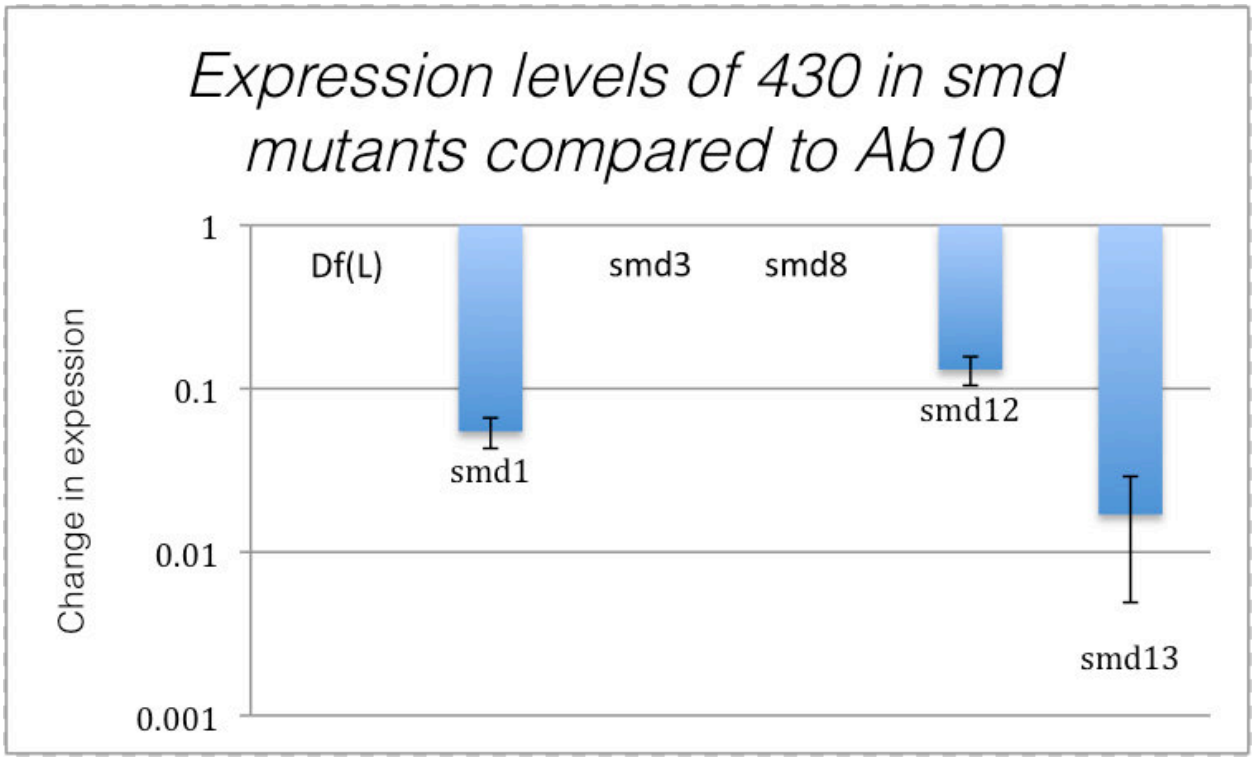

Figure 2.4: Gene 430 is downregulated in homozygous *smd1/smd1*, *smd12/smd12*, and *smd13/13* as compared to progenitor Ab10. Gene 430 is genomically absent in Df(L), *smd3*, and *smd8*. This work was done in collaboration with Amy Webster as a part of her undergraduate thesis.

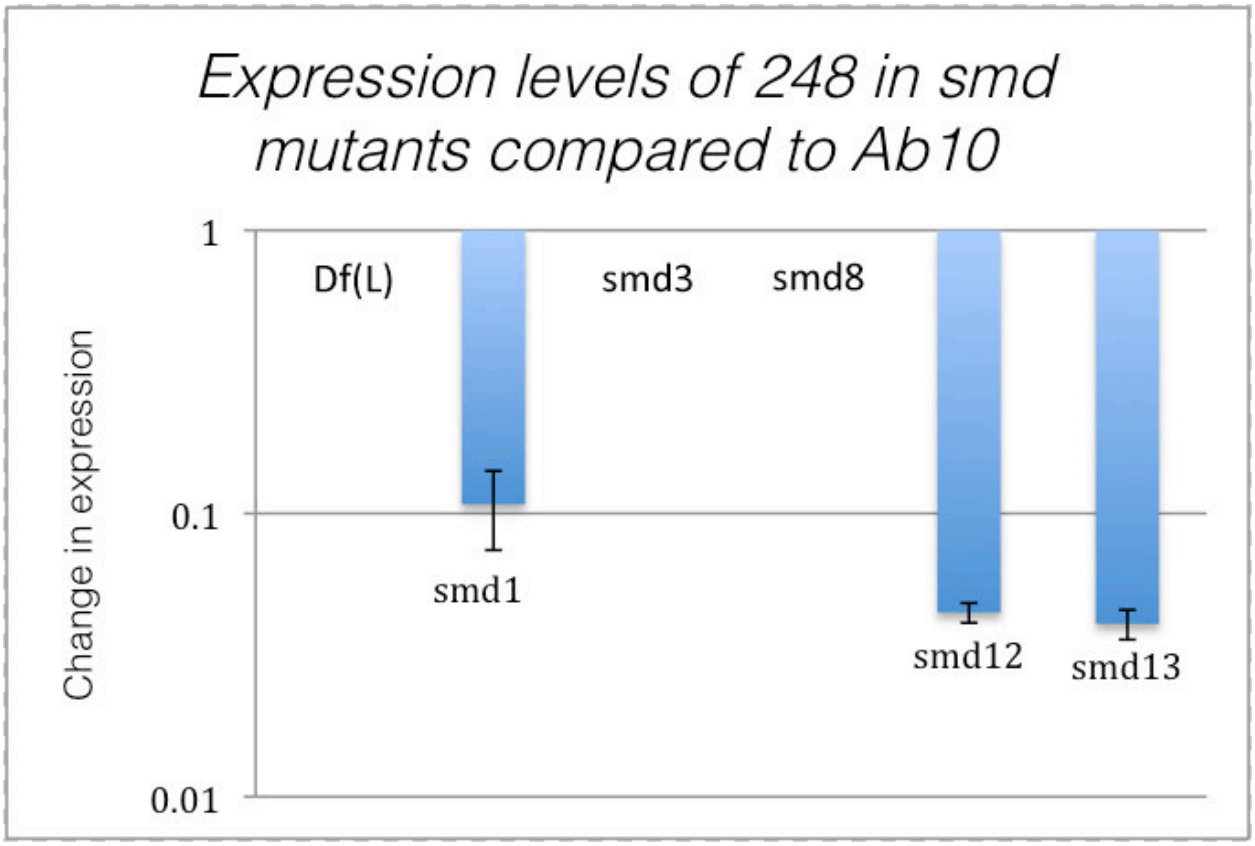

Figure 2.5: Gene 248 is downregulated in homozygous *smd1/smd1*, *smd12/smd12*, and *smd13/13* as compared to progenitor Ab10. Gene 248 is genomically absent in Df(L), *smd3*, and *smd8*. This work was done in collaboration with Amy Webster as a part of her undergraduate thesis.

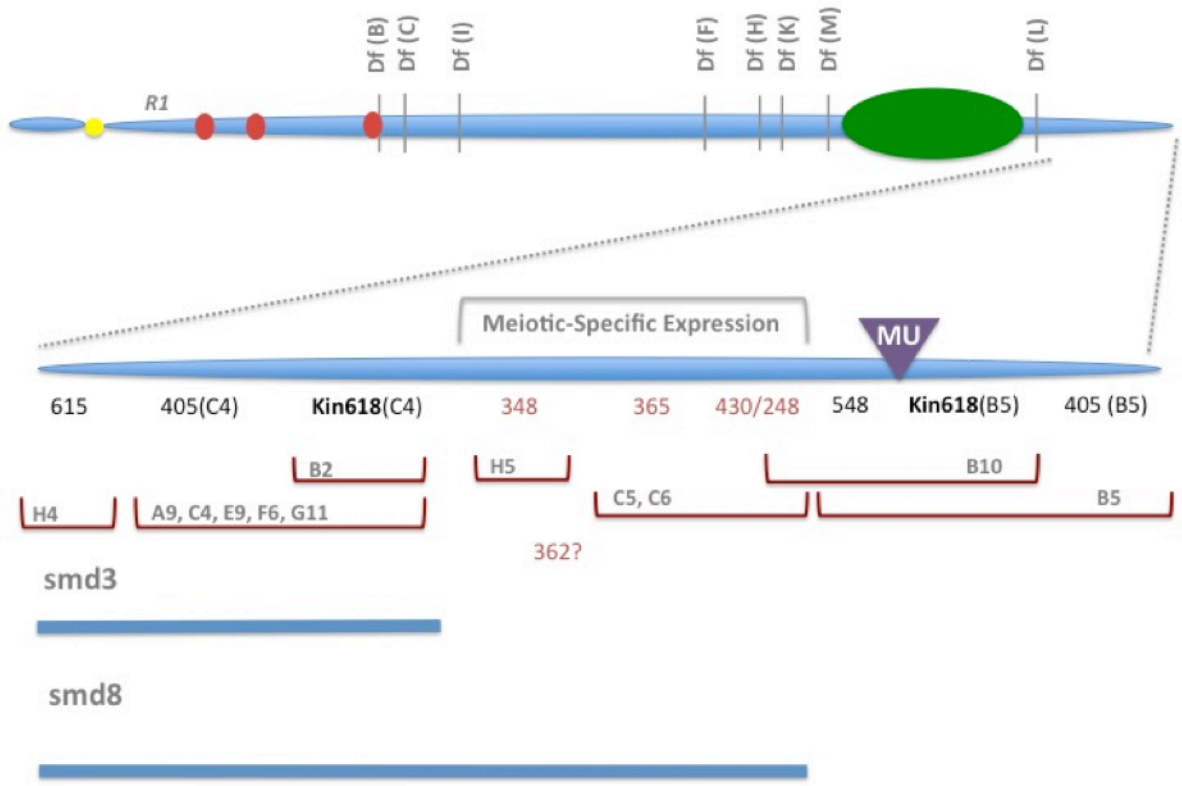

Figure 2.6: Genetic Map of the distal tip of Ab10 showing the relative positions of the 9 distal transcripts. Yellow marks the centromere, red marks the three TR-1 knobs and green marks Knob 180. Map is not to scale, with long arm of Ab10 enlarged for clarity. Deficiency lines are shown in grey, with the Df(L) breakpoint shown right before the distal tip. BACs are represented by red brackets. The Kin618 (B5) *Mu* element is represented by the purple triangle (See Chapter 3). Distal genes only expressed in meiotic tissue are in red. The locations of the genes are approximate and relative, based on the *smd* deletions and BAC content (see text). Both gene 405 and Kin618 likely have at least two copies (more evidence for this is presented in Chapter 3). The genes only expressed during meiosis (348, 365, 430, and 248) are all in the same region on the distal tip. Since the BAC probe for 362 did not work, its exact placement on the distal tip is unclear but it is distal to the *smd3* breakpoint as assayed by PCR.

CHAPTER 3: THE SMD12 EPIMUTATION SILENCES A NOVEL KINESIN GENE FAMILY INVOLVED IN MEIOTIC DRIVE OF ABNORMAL CHROMOSOME 10

Abstract

The preferential segregation of Abnormal Chromosome 10 in *Zea mays* is characterized by the unusual activity of knobs, or balled regions of heterochromatin, during meiosis. Instead of allowing centromeres to pull chromosomes towards opposite ends of the dividing meiotic cell, knobs on Ab10 form neocentromeres and move independently along microtubules during Anaphase. Here we report the discovery of Kin618, a C-terminal kinesin specific to Ab10 involved in Knob 180 neocentromere movement. A mutant Ab10 deficient in meiotic drive, *smd12*, does not transcribe Kin618, have Knob 180 neocentromeres, nor preferentially segregate. Kin618 is multi-copy and *smd12* is an epimutation in which all active copies of Kin618 have been silenced through DNA methylation.

Introduction

The genome of *Zea mays* is punctuated by large, dense regions of heterochromatin called knobs. Most maize lines contain between 4 and 8 knobs at any of approximately 33 locations within the genome [1,2]. Knobs are composed of tandem repeats of two sequences: TR-1, a 350 base pair sequence and Knob 180, a 180 base pair sequence [3,4]. Though some knobs are

composed solely of the TR-1 repeat, most knobs contain only Knob 180 or a combination of both [5]. Knobs are unremarkable except in the presence of Abnormal Chromosome 10, a haplotype variant of the tenth maize chromosome. Without Ab10, meiotic division proceeds canonically: microtubule spindles attach end-on to proteinaceous kinetochores at the centomere and pull the chromosomes to opposite ends of the dividing cell [6]. However, in the presence of Ab10, knobs are activated into neocentromeres that slide laterally along the microtubule spindles [7,8]. Neocentromeres move much faster than centromeres, and the knobs streak ahead to the poles of the dividing cell, dragging the rest of the chromosome behind [8].

This dramatic movement results in preferential segregation, as the knobbed chromosomes meiotically drive themselves into the cell destined to develop into the egg [9]. A model that links neocentromere activity and meiotic drive was developed by Marcus Rhoades who pointed out that in maize, the bottom cell of female meiosis will be fertilized by the pollen and develop into the seed [10]. Neocentromeres pull knobs poleward and into that predestined cell. Neocentromere activity can cause Ab10 to segregate at levels up to 83% through the female [11].

The molecular mechanism of knob movement is mysterious. The spindle microtubule is composed of polarized tubulin that is constantly assembling and disassembling in flux [12]. Canonical chromosome movement is a carefully orchestrated process involving specialized kinetochore proteins and microtubule flux [13]. Neocentromeres are entirely different from centromeres and do not have the histone variant CenH3 nor any of the known proteins associated with kinetochores [14–16]. They move laterally along the microtubule spindle toward the poles and, unlike centromeres, are unaffected by tubulin flux [17]. Previous observational studies have proposed that an Ab10-specific kinesin, a molecular motor specific to microtubules, walks along the spindle and pulls neocentromeres to the poles [17].

Kinesins are molecular motor proteins that "walk" on microtubules and perform a wide range of cellular functions, from transporting organelles to orchestrating chromosome movement during cell division [18,19]. They are characterized by a catalytic "motor" domain, which contacts the microtubules and hydrolyzes ATP, allowing movement [20]. The "tail" domain of the kinesin is functionally specific and binds to the appropriate cargo. The position of the motor domain within the kinesin determines the direction it walks along the polarized microtubules: a catalytic core at the N-terminus confers plus-end directed motility while kinesins with the motor domain at the C-terminus move towards the minus-end of the microtubules. Molecular phylogenetics places kinesins into 14 distinct families [21,22]. Within families, individual kinesins vary in whether they function as monomers, hetero/homodimers, heterotrimers, or hetero/homotetramers [23–25].

Though both TR-1 and Knob 180 form neocentromeres in the presence of Ab10, they are controlled by separate trans-acting factors. The long arm of Ab10 is larger than normal chromosome 10 and has three small TR-1 knobs, a large Knob 180, and a "distal tip" domain [26]. A series of deficiency lines of Ab10-I with breakpoints arrayed along the long arm of the chromosome have allowed mapping of integral functions (Df(B), Df(C), Df(I), Df(F), Df(K), Df(H), Df(M), and Df(L)) [26]. The smallest truncation, Df(L), lacks the distal tip and does not have Knob 180 neocentromere activity nor the ability to preferentially segregate [17,27]. TR-1 neocentromeres are still activated when Df(L) is present in the plant, indicating that there are two distinct neocentromere motors that are sequence specific [27]. It is the Knob 180 neocentromere motor that specifically correlates with meiotic drive.

The Dawe lab has five suppressor of meiotic drive (*smd*) mutants in which the Abnormal chromosome no longer preferentially segregates [26,28]. The *smd* mutants were created by transposon mutagenesis: a line active for the DNA transposon Robertson's Mutator (*Mu*) was

crossed to an Ab10 line and the progeny were screened for lack of Ab10 drive (50% segregation). Two of the five mutant lines, *smd3* and *smd8*, are terminal deletions (Chapter 2). Though much effort has been expended to localize the *Mu* element causing the phenotype in *smd1*, *smd12*, and *smd13*, including extensive *Mu* Illumina sequencing, a novel *Mu* insertion unique to those lines has still not been found [29].

Creating a *Mu* active line is an inaccurate process that may lead to structural or epigenetic changes in the maize genome. *Mu* transposons themselves are activated or inactivated through DNA methylation in their terminal inverted repeats (TIRs) [30,31]. In fact in plants, the majority of DNA methylation is found on transposons or repetitive elements [32]. There are three motifs in plants that become methylated: CG, CHG, and CHH (where H is any nucleotide except G). Though all types of plant methylation are established by RNA-directed DNA methylation (RdDM), they are maintained by different pathways and may perform distinct functions [33,34]. In maize, methylation at CG and CHG is associated with deep heterochromatin and transcriptional inactivation [35]. By contrast, transcriptionally active regions have low levels of CG and CHG methylation. More elusive is the function of the asymmetric CHH methylation, though in maize it has been revealed to be particularly associated with transposons within 1 kb of maize genes [35]. Though CHH methylation has previously been thought to recruit CG and CHG methylation to form silenced heterochromatin, its distribution pattern within the maize genome suggests a more nuanced role.

The effect of methylation on gene transcription allows for a new level of mutational possibility. As the presence of methylation at CG or CHG motifs within a gene prevents transcription of the gene, changes in methylation patterns across genes can lead to expression changes known as epimutations. Epimutants are rare, though a number have been characterized

specifically in plants. These epialleles, as they are also called, have been found to alter floral morphology, gene expression, and disease resistance in *Arabidopsis thaliana*, pigmentation in *Zea mays*, and floral symmetry in *Linaria vulgaris* [36–40]. Plants may be more susceptible to the development of epialleles due to the lack of division between germ line and soma, making them less efficient in resetting methylation between generations [41]. Plants are also rife with transposable elements, and some attribute the origin of DNA methylation to the plant's defense against transposons [42]. Indeed, mutations in maize methylation machinery have led to widespread upregulation of DNA transposons, including *Mutator* [43]. The characterization of a naturally occurring *Mu* killer (*Muk*) locus, which specifically silences only *Mu* transposons, also supports the view of defensive DNA methylation [44]. *Muk* is the product of an inverted duplication of a truncated *Mutator* transposon and produces a hairpin transcript that induces silencing of all genomic *Mu* elements. In the case of *Muk*, a small alteration in the DNA sequence has resulted in widespread epigenetic changes. Further study of epimutations, will refine our understanding of the relationship between sequence alterations, epigenetics, transposons, and gene expression in plants.

This study presents the discovery of a C-terminal kinesin specific to Abnormal Chromosome 10, which we call Kin618. There are multiple copies of Kin618 on the "distal tip" of Ab10 that vary in promoter sequence as well as a series of SNPs throughout the coding region. Characterization of the epimutant *smd12*, in which all copies of Kin618 are silenced by DNA methylation, links Kin618 to Knob 180 neocentromere movement. Kin618 retains closest homology to maize kinesin 11, and the research described here presents a clearer picture of the origin and molecular mechanism of the Abnormal Chromosome 10 meiotic drive system in *Zea mays*.

Materials and Methods

Annotation and Homology of Kin618

A de novo transcriptome assembly specific to transcripts unique to the distal region of Ab10 originally identified the Kin618 RNA transcript (Chapter 2). NCBI BLASTX identified the transcript as a C-terminal kinesin [45]. NCBI BLAST to the nonredundant nucleotide database identified the closest maize homolog at the nucleotide level. An open reading frame and the translation start site were identified using the Colorado State Molecular Toolkit translation tool (http://www.vivo.colostate.edu/molkit/index.html). Homology between Kin618 and Kinesin 11 was quantified using ClustalW2 [46].

Specificity of Kin618 to the Meiotic Drive Phenotype

Primers were designed using Primer3 to the Kin618 RNA transcript and tested on leaf tissue genomic DNA from a homozygous Df(L) line, which features Ab10-I with a truncated "distal tip" area, and a homozygous Ab10-I line to test presence on the distal tip and effectiveness of primers [47]. Further analysis located these primers in the fourth and fifth exon of the Kin618 transcript and they will be referred to as *Kin618 (all).* The position of Kin618 was further confirmed by PCR across a series of deficiency lines that are versions of Ab10-I that have been truncated at various positions along the long arm of the chromosome: Df(B), Df(C), Df(I), Df(F), Df(K), Df(H), and Df(M). DNA was extracted using CTAB and PCR was performed using Phusion polymerase mastermix as described in Chapter 2. Products were visualized on a 2% agarose gel. Using the same protocol, DNA was also extracted from 2 week seedlings heterozygous for Ab10-II, Ab10-III and K10L2 and PCR was performed using B73 as a control

to check the presence of Kin618. PCR was also performed across 47 inbred lines containing (~55%) of the diversity present in maize.

Kin618 BAC Preparation, Illumina Sequencing, and Assembly

Overgo probes specific to Kin618 identified 24 BACs containing the kinesin in the Ab10 homozygous 4X BAC library, as presented in Chapter 2. Eight of these 24 BACs had already been sequenced, so the remaining 16 BACs with Kin618 were prepared for sequencing. Frozen culture from each BAC was scraped and plated on LB Agar plates containing 12.5 μg/mL chloramphenicol and allowed to grow at 37 °C overnight. Colonies were picked and used to inoculate 5 mL LB broth with 12.5 μg/mL chloramphenical and grown for approximately 7 hours. One mL of the 5 mL culture was subsequently used to inoculate a 500 mL culture and grown for approximately 16 hours. Cells were pelleted at 6,000 rpm for 25 minutes and BAC DNA was extracted using the Qiagen Large Construct Kit. BAC DNA was sent to GGF for library creation with TruSeq DNA and MiSeq PE300 Illumina sequencing. Reads were trimmed and cleaned using FastX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Contigs were assembled using Velvet Optimiser and contigs containing Kin618 were identified using NCBI blast [48].

PacBio Sequencing and Assembly of Kin618-Containing BAC clones

Two BAC clones, C4 and B5, were targeted for assembly of the entire Kin618 gene. BACs were grown and purified as above but, since PacBio sequencing requires a large amount of starting material, 6 preps were done for each BAC (C4 and B5) and DNA was combined into a single sample. DNA was sent to the UC Davis Sequencing Core, which made 5 kb insert libraries

for SMRT sequencing. PacBio reads in FASTA files were error-corrected using the complementing BAC Illumina reads (Table 2.1) by PacBiotoCA and a first pass assembly was made using the Celera Assembler [49]. NCBI BLAST pulled out assembled contigs containing the Kin618 gene, which were further assembled using Sequencher (http://www.genecodes.com/). Kin618 gene annotation was performed by TopHat [50] to map Ab10 RNA-seq Illumina reads to the C4 and B5 genome contigs. Alignment and calculation of similarity scores between C4 and B5 RNA and protein transcripts was performed by Geneious [51].

The E7 and B2 BACs were prepared for PacBio sequencing as described above. DNA was sent to the ICBR Nextgen DNA Sequencing core at the University of Florida for library creation and SMRT sequencing. PacBio fasta files were error-corrected using LSC and the BAC Illumina reads described previously (Table 2.1)[52]. Corrected PacBio reads containing Kin618 were identified using NCBI BLAST and extracted via a Perl script for contig assembly using Sequencher. The contigs from E7 did not assemble into any clear gene model, and it is unclear if this is due to poor sequence quality or because E7 does not have a complete Kin618 gene. The B2 Kin618 copy was identified using NCBI BLAST of the B5 Kin618 cDNA to the 7 largest contigs of the B2 BAC assembly. Annotation of the B2 Kin618 gene was attempted using TopHat to map Ab10 RNA-seq Illumina reads (GGF_08_SOL_0001_TGACCA.fq) back to the B2 genomic contig, however was unsuccesful (Table 3.1). According to TopHat there is no expressed Kin618 copy from the B2 genomic contig.

Conservation of Both Copies of Kin618

Seeds were germinated from four teosinte and two landrace lines containing the abnormal chromosome (Table 3.3). DNA was extracted from 2 week old seedlings using CTAB as

described in Chapter 2. Primers were designed to the promoter region of both the C4 and B5 copy using Primer3 software that identified each copy specifically: *Kin618 (B5)* and *Kin618 (C4)* (Table 3.2). PCR was performed as described in Chapter 2. Products were visualized on a 2% agarose gel.

Kin618 Expression Assay

Immature tassels were collected from homozygous mutant lines *smd1/1*, *smd3/3*, *smd8/8*, *smd12/12*, *smd13/13* as well as homozygous Df(L) and homozygous Ab10-I. For each genotype, tissue from two sibling plants (biological replicates) was collected. Anthers were removed from florets into 1X PBS. Meiocytes were extruded from anthers into PBS and visualized to ensure proper staging of meiosis. Anthers containing cells between prophase I to telophase II were flash frozen using liquid nitrogen. RNA was later extracted using the Qiagen RNeasy Kit. Frozen anthers were vortexed and ground with an electric pestle in Buffer RF of the Qiagen kit and the protocol was followed as specified in the kit. Equal amounts of RNA from each sample was converted to cDNA using SuperScript® III First Strand Synthesis Kit with Poly(A) specific primers. Presence of Kin618 was scored using *Kin618 (all)* primers. B-tubulin primers acted as a positive control. PCR was performed on the cDNA as described above. Reactions were run on a 2% agarose gel to check presence/absence of expression. The study was repeated several months later with plants grown separately from the original samples. The second assay scored two homozygous Ab10-I plants, two homozygous Df(L) plants, two *smd8/smd8* plants and two more *smd12/smd12* plants, for a total of four *smd12/smd12* plants scored in total.

To further verify the results, qRT-PCR was performed across all genotypes (Ab10I, Df(L), *smd1*, *smd3*, *smd8*, *smd12*, *smd13*), with two biological replicates each on the first round

of samples extracted. Reactions were performed using SYBRGreen mastermix with an internal control of β-actin by Amy Webster. Standard dilution curves were calculated for each primer to insure efficiency. Each reaction was performed in triplicate and fold change was calculated using ΔΔCt relative to expression in wild type Ab10 lines.

FISH of smd12 mutant and Df(L) deficiency line

Immature tassels were extracted and staged as described previously for plants heterozygous for both smd12/N10, Df(L)/N10 and Ab10/N10. Anthers were fixed in 4% paraformaldehyde and coverslips were prepared as described previously [53]. FISH for TR-1, Cent-C, and Knob 180 was performed as described previously [5]. Slides were imaged using the Zeiss Axio Imager and processed using Slidebook 5.0 (Intelligent Imaging Innovations, Denver, CO, USA).

Targeted Bisulfite Sequencing of Multiple Kin618 Copies

Homozygous *smd12* and Ab10-I genomic DNA was collected from 2 week old seedlings and bisulfite treated using the EpiTect Bisulfite Kit from Qiagen. Four sets of degenerate primers were designed using Kismeth directly upstream of both the C4 and B5 translation start site, within the first intron of the B5 genomic copy, and spanning the TIR/ gene-specific boundary of the *Mu* element directly upstream of the B5 genomic copy (Table 3.2). All primers are specific to either the B5 or C4 copy. PCR was performed on the bisulfite-treated DNA using Zymotaq enzyme which is optimized for bisulfite treated DNA. PCR products were gel-purified and cloned in E.coli using the TOPO10 cloning kit and grown on LB agar plates containing ampicillin. Between 7 and 13 single colonies were picked for each section and colony PCR was

performed using the M13 primers. PCR products were sent to Georgia Genomics Facility (GGF) for Sanger Sequencing. Results were visualized using the Kismeth software [54].

Identification of 6 Unique Copies of Kin618

Unassembled Illumina reads from all 24 BACS were mapped back to the B5 PacBio assembled contig using Bowtie2 [55]. SNPs between them were called and visualized using Integrated Genomics Viewer (IGV) [56] and organized into 6 unique profiles by eye.

Results

Discovery of an Ab10-specific Kinesin

Among the genes identified using the de novo transcriptome comparison approach (Chapter 2), the most promising was a C-terminal kinesin we call Kin618. Using deficiency line mapping, the genomic presence of Kin618 was confirmed to be on the distal tip of the abnormal chromosome and nowhere else in the maize genome (Figure 3.1). The Kin618 gene has closest homology to Kinesin 11, a C-terminal kinesin on maize chromosome 7 whose Arabidopsis homolog, Atk5, is a known participant in spindle assembly[57,58]. Kin618 contains a truncated N-terminal cargo domain as compared to maize Kinesin 11 (Figure 3.3). The conserved region between Kin11 and Kin618 retains 88% similarity at the nucleotide level and spans the C-terminal motor domain. The unique N-terminus of Kin618 is composed of a single first exon that features a series of highly charged amino acids, which hints at a possible DNA-binding function. The sequence of the first exon is key to the functional divergence between Kin618 and Kin11 that allows for a specialized drive function.

Kin618 is both unique and specific to the drive phenotype. Of the four chromosome 10 haplotype variants, only those that preferentially segregate contain Kin618 (Ab10-I, Ab10-II, and Ab10-III) (Figure 3.4). K10L2, which shows TR-1 neocentromere activity but lacks Knob 180 activity and does not preferentially segregate, lacks Kin618. The presence of Kin618 was assayed using PCR across 47 inbred lines without Ab10 that represent the majority of the diversity of maize. All lacked Kin618, indicating that the presence of the kinesin is specific to the abnormal haplotype.

Kin618 is a Multicopy Gene Family

Kin618 was discovered in a de novo transcriptome assembly which, due to the abundance of isoforms and the homology between unique maize kinesins, proved unreliable in the development of an accurate gene model. To properly annotate Kin618, we turned to a 4X homozygous Ab10 BAC library which, when sequenced and assembled, could provide us with full genomic sequences of individual members of the Kin618 gene family. Overgo probe hybridization revealed that 24 BACs contained Kin618. Of these 24, we had sequenced 8 for a previous study (Chapter 2). To obtain all genomic sequence of Kin618 in the BAC library, we sequenced the remaining 16 BACs using Illumina technology (Table 3.1). We then attempted a de novo assembly of the Kin618 gene space using sequence from all 23 BACs (one of the 24 BACs failed in sequencing). Unfortunately, the highly repetitive nature of the BACs precluded assembly of the full Kin618 gene from any of the BACs sequenced by Illumina technology. However, from fragmented assemblies, it appeared that several copies of Kin618 were represented among the BACs.

We picked three BAC clones that appear to have different Kin618 promoters, B5, C4, and B2, and assembled them by PacBio sequencing. The most successful assembly was for BAC B5, which yielded a 40 kb contig with Kin618. It also revealed a native *Mutator* transposable element approximately 200 bases upstream of the transcription start site of B5 Kin618. The C4 PacBio assembly was partially successful and resulted in three long contigs: one containing the first exon, one containing the second exon, and one containing the remaining body exons. The B2 assembly yielded a 22 kb contig that, according to BLAST with the Kin618 cDNA, contains Kin618 (Figure 3.2, 3.3). Therefore the PacBio sequencing confirmed the presence of at least three unique copies of Kin618 on the distal tip of Ab10 (Figure 3.5).

To quantify Kin618 gene copy number, we aligned Illumina genomic reads from all 23 BACS to the B5 PacBio genomic assembly. By identifying SNPs between B5 reads and reads of each of the other 23 BACs, we were able to create 6 SNP profiles to which 20 of the BACs clearly sorted (Figure 3.6). Three of the BACS had unclear SNP profiles. The six distinct SNP profiles suggest a six-member gene family. This hypothesis finds support from the 24 BACs containing Kin618 in a library with only 4X coverage.

B5, C4, and B2 Diverge in Sequence

The genomic sequence of B5, C4, and B2 allowed a comparison of promoters between the two fully assembled Kin618 genes and B2 (Figure 3.7). The B5 and C4 promoter are highly similar for approximately 250 base pairs upstream of the translation start site, save for a 10 nucleotide deletion in the C4 promoter. The B2 promoter, by contrast, has many SNPs and indels relative to the other two genes, including an insertion in the “ATG” start codon. The B5 and C4

promoter begin to diverge upstream of the coding sequence where the B5 copy has a *Mutator* transposable element that both the C4 and B2 copy lack.

To annotate the three Kin618 genes, we used TopHat to map Ab10 Illumina reads of the anther transcriptome back to the genomic assembly of each gene copy. According to TopHat, there is no expressed gene from the B2 contig. Patches of transcriptome reads that did align to the B2 contig had many SNPs and it appears that the remnants of the Kin618 exons within the B2 contig have many base pair changes. It is very likely that the B2 contig presents a copy of Kin618 that is no longer expressed or functional. By contrast, the annotation yielded a clear gene model for both the C4 and B5 copy of Kin618, C4 with 16 exons and B5 with 15 exons. However, the distribution of the exons varied between the two copies (Figure 3.3). B5 and C4 are identical in the region that they share with Kin11, from midway in the second exon of Kin618 to midway through the 12th exon, the region that spans the C-terminal motor domain. The three exons of B5 and the four exons of C4 distal to the region of Kin11 homology are all noncoding. They are also identical, though C4 appears to have an extra exon and an additional 584 nucleotides.

The crucial variation between the C4 and B5 copy is in the first exon of Kin618. A very clear isoform appears in the B5 copy in which an intron appears within the first exon whereas this is not seen in the C4 copy. When the first exon is unspliced, as it is in C4, it is a tandem repeat. However, the spliced B5 version removes the repeat (Figure 3.3).

Interestingly, the coding region of B5 and C4 are identical except for 8 SNPS that differentiate the two genes. These SNPs distinguishing both C4 and B5 are present in the transcriptome data, maintaining the hypothesis that both copies are expressed during meiosis. The SNPS between the two copies result in 8 amino acid changes within the first 400 amino

acids of the predicted protein sequence (B5 considered with an unspliced first exon) (Figure 3.8). Seven of the eight changes are in the N terminal tail domain, hinting at a functional variation between Kin618 C4 and Kin618 B5.

Multiple Kin618 Copies May Be Necessary

To determine if more than one expressed copy of Kin618 is necessary for the drive function, we used B5 and C4 promoter-specific primers to check the presence of both copies across a diverse panel of teosintes and landraces genotyped as Ab10 positive. Our samples included both *mexicana* and *parviglumis* and landraces originating from Mexico and Brazil (Table 3.3). Both B5 and C4 are conserved across both subspecies and geographic boundaries (Figure 3.9). Furthermore, the deletion mutations *smd8* and *smd3* (Chapter 2) contain the C4 Kin618 copy but lack the B5 Kin618 copy (Figure 3.10). Mutation *smd3* particularly is a small deletion and may point to the importance of the B5 copy.

Kin618 is not expressed in meiotic drive mutant smd12

Though the gene Kin618 seems the ideal candidate for moving neocentromeres to the poles of a meiotic cell, a strong genetic or molecular link is needed to confirm this hypothesized function. Both *smd3* and *smd8* were revealed to be deletion mutations, but *smd1*, *smd12*, and *smd13* remained uncharacterized. As all the *smd* lines were isolated from a Robertson's *Mu* screen, the cause of the mutations may be due to *Mu* insertions into the gene(s) on Ab10 that causes drive. However, not all mutants identified in *Mu*-active lines are caused by *Mu*.

If Kin618 causes meiotic drive, we would expect to see a difference between mutant and wild-type Ab10 plants in the expression level of this gene. Extracted RNA from all five mutant

meiocytes was assayed for expression levels of Kin618 using primers specific to a region shared across all 6 copies (*Kin618 all*). Expression of Kin618 was also assayed in both progenitor Ab10 and the deficiency line Df(L) which lacks all genomic copies of Kin618 and is a negative control. For each genotype, expression was measured in two sibling plants, providing a biological replicate.

Two plants of one of the mutants, *smd12*, lacked expression of Kin618 (Figure 3.11). The experiment was replicated two months later with two more *smd12* plants as well as two more Ab10-I, Df(L) and *smd8* plants grown under an independent growth cycle, and the results confirmed. Downregulation of Kin618 in *smd12* was further quantified using qRT-PCR and compared across all *smd* mutants (Figure 3.12). As can be seen, the expression of Kin618 in *smd12* is comparable to the levels observed in the Ab10 deletion line Df(L), which lacks the entire Kin618 gene family. These data suggest that the *smd12* mutation affects all copies of the Kin618 gene family and reduces expression to nearly zero.

Knob 180 neocentromeres are inactive in smd12

Rhoades's model postulates that neocentromere movement results in preferential segregation of the abnormal chromosome. The *smd12* mutant does not preferentially segregate, therefore we used FISH on meiocytes of *smd12* plants to look at neocentromere movement during meiotic anaphase. Knob 180 does not form neocentromeres in *smd12*, however, TR-1 neocentromeres are still active (Figure 3.13). The same phenotype is observed in the Ab10 deletion line Df(L)[27]. This meiotic phenotype offers a strong link between Kin618 expression, Knob 180 movement, and meiotic drive.

Smd12 is an epimutation

Since Kin618 appears to be a gene family with at least two copies, C4 and B5, actively expressed, the complete absence of Kin618 transcript in the *smd12* mutant was puzzling. This global silencing effect prompted us to examine the epigenetic state of both C4 and B5 in *smd12*. To test for the presence of DNA methylation, we bisulfite treated Ab10 and *smd12* DNA and designed degenerate primers to amplify a 198 bp GC-rich segment directly upstream of the translation start site of B5 Kin618, approximately 325 bases downstream of the 3' terminal inverted repeat (TIR) of the *Mu* element. These primers are specific to B5 and do not target C4 nor B2 (Figure 3.7). Excitingly, 22 CG sites and 6 CHG sites within the 198 bp segment were methylated in over half of the clones of the bisulfite-treated *smd12* samples that were not methylated in the Ab10 control (Figure 3.14). We extended this analysis into the homologous 188 bp segment in front of the translation start site of C4 Kin618 and found an increase in mutant DNA methylation at 23 CG sites and 3 CHG sites in over half of the sequenced clones (Figure 3.15). The C4 primers are also completely specific to the C4 copy (Figure 3.7). Targeted bisulfite sequencing of a 215 bp segment of the first intron of B5 Kin618, revealed an increase in CHH methylation at 20 sites and CG methylation at 1 site in the mutant (Figure 3.16). The intron primer pair is specific to B5 and will not amplify either C4 nor B2, since C4 has a 72 nucleotide insertion that the forward primer spans and B2 has a 76 base pair insertion that the forward primer spans. Finally, we performed targeted bisulfite sequencing using gene-specific and TIR-specific primers to assay any change in methylation state of the proximal TIR of the *Mu* element directly upstream of B5 Kin618. Our results point to a decrease in CHH methylation at 30 sites in the *smd12* mutant as compared to the wildtype Ab10 (Figure 3.17).

Discussion

The molecular mechanism of neocentromere movement is an elusive topic. The lateral sliding action of neocentromeres along the meiotic spindle suggests a microtubule-specific motor protein such as a dynein or a kinesin that attaches to knobs and guides them towards the poles. Since there are no known dyneins in maize, a kinesin is the best candidate [59]. Biologists divide kinesins into fourteen families that vary in their directionality of movement on the polarized microtubules [21]. Kin618 belongs to the minus-end directed Kinesin-14a family that includes Ncd in Drosophila, KIFC1/KIFC5a in mouse, and Atk5 in Arabidopsis[22,57,60]. Within the species, Kin618 most resembles Kinesin 11, the maize homolog of Atk5. KifC5a, Ncd, and Atk5 all play roles in spindle formation and proper chromosome segregation in meiosis and mitosis [57,58,61–63]. Atk5, the closest homolog of Kin618, binds microtubules with its N-terminal tail and bundles them together to form the bipolar spindle [57,58].

The tail domain of Kin618 has diverged from this function. Comparison between Kin11 and Kin618 shows that the 5 exons of Kin11 that encode the microtubule bundling domain of that kinesin are missing in Kin618 (Figure 3.3). Instead, Kin618 contains either a single exon or, in the case of the spliced B5 isoform, two small exons. This shortened tail featured a series of highly charged amino acids that may bind directly to DNA, presumable to sequence-specific Knob 180 repeats. When unspliced, the first exon of Kin618 features a tandem duplication. If the role of the protein product of the first exon is to bind to DNA, this repeat may increase the efficiency of binding. The tandem repeat nature of Knob 180 may be reflected in the tandem repeat nature of the protein that facilitates neocentromere formation. Perhaps this unique feature evolved after duplication of maize Kinesin 11 during the last paleotetraploidy event in maize.

One of the copies was co-opted for a new, more selfish function. Once Kin618 became genetically linked to Knob-180, Ab10 began to meiotically drive.

By harnessing the diversity present within maize, we show that Kin618 is unique and specific to the Ab10 chromosome. Unlike maize Kinesin 11, multiple copies of Kin618 exist on the "distal tip" of Ab10. SNP profiling suggests at least six copies, though so far only three are confirmed: B5, C4, and B2. One of the three copies (B2) appears to not be expressed and does not present us with a clear annotated gene model. Conversely, the B5 and C4 copies are likely transcribed. The B5 and C4 predicted cDNAs vary by 8 SNPs in the N-terminal tail region. These SNPs result in predicted changes at the amino acid level, though this study cannot confirm that either gene is translated. The B5 copy also features variation in an alternatively spliced first exon, which may function differently than unspliced C4. However, since both B5 and C4 are conserved across subspecies and geographic boundaries it is likely that both are translated and crucial for the drive phenotype. Kin618 may function as a heterodimer, or there may be several isoforms that perform crucial functions. Characterization of the three remaining copies of Kin618 will further clarify the role of Kin618 in neocentromere movement.

The genetic link between Kin618 and drive comes from the epimutant *smd12,* which lacks transcription of Kin618 and does not form Knob 180 neocentromeres nor preferentially segregate. Epimutations are rare. The few known in plants affect morphology, seed pigmentation, pathogen resistance, and development [40,64]. In the *smd12* mutation, targeted bisulfite sequencing reveals significant increases in CG and CHG methylation in the region directly upstream of the translation start site of both the C4 and B5 copies of Kin618. Interestingly, we also see an increase in CHH methylation in the first intron of B5, suggesting continuous RdDM in that region. A decrease in CHH methylation is also seen in the TIR of the

Mu transposable element directly upstream of Kin618 (B5). B5 and C4 have high homology across the promoter and first exon so sequence-targeted DNA methylation would silence all copies of Kin618. Perhaps RdDM was induced by a small structural change, such as an inversion, near one copy of Kin618 that then homologously silenced the other five. Studies of epialleles in Arabidopsis have shown that small sequence deletions or rearrangements can induce epigenetic changes at more distant genomic regions, changing genic expression levels [65]. Two of the five *smd* mutants isolated from the Robertson's *Mu* screen are terminal deletions, so structural changes are not uncommon when using *Mutator*.

Though *smd12* may very likely be the result of a structurally induced epigenetic change, the presence of a native *Mu* transposable element within 200 bases of the transcriptional start site of B5 gives one pause. Since *smd12* is the descendent of a parental genotype with active *Mu* transposons, it is attractive to speculate that a change in the epigenetic state of the native *Mu* spread into the B5 gene and silenced its expression. Active *Mu* lines have been known to spontaneously silence themselves via epigenetic changes after several generations of inbreeding [30,66]. In maize particularly, heterochromatin in retrotransposons has been shown to spread into neighboring genes [67]. Other studies suggest this spreading leads to maize epialleles [68]. Perhaps the active *Mu smd12* line spontaneously silenced, an event that epigenetically altered even native *Mu* elements, such as the one upstream of B5. The silencing effect may then have spread into the proximal B5 Kin618 gene, methylating the other Kin618 copies due to sequence homology. However, other analyses point to increases in CG and CHG methylation in *Mu* TIRs, as opposed to CHH methylation, as crucial to *Mu* epigenetic silencing (D. Lisch personal correspondence). Mutant *smd12* only features a change in CHH methylation specific to the *Mu*. Nor de we fully understand the relationship between CHH, CG, and CHG methylation in maize. Further

characterization of the *smd12* epimutation has great potential to clarify the careful balance between epigenetics and gene expression within the maize genome.

Acknowledgements

We would like to thank Amy Webster for conducting the qRT-PCR experiments.

References

1. Albert, P. S., Gao, Z., Danilova, T. V & Birchler, J. A. Diversity of chromosomal karyotypes in maize and its relatives. Cytogenetic and genome research 129, 6–16 (2010).
2. Kato, Y. Cytological studies of maize (Zea mays L.) and teosinte (Zea mexicana Shrader Kuntze) in relation to their orIgin and evolution. Mass Agric Exp Sta Bull 635, 1–185 (1976).
3. Peacock, W. J., Dennis, E. S., Rhoades, M. M. & Pryor, A. J. Highly repeated DNA sequence limited to knob heterochromatin in maize. Proceedings of the National Academy of Sciences of the United States of America 78, 4490–4 (1981).
4. Ananiev, E. V, Phillips, R. L. & Rines, H. W. A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? Proceedings of the National Academy of Sciences of the United States of America 95, 10785–90 (1998).
5. Kanizay, L. B., Albert, P. S., Birchler, J. A. & Dawe, R. K. Intragenomic conflict between the two major knob repeats of maize. Genetics 194, 81–9 (2013).
6. Cheeseman, I. M. & Desai, A. Molecular architecture of the kinetochore-microtubule interface. Nature reviews. Molecular cell biology 9, 33–46 (2008).
7. Rhoades, M. Preferential segregation in maize. Heterosis 66–80 (1952).
8. Yu, H. G., Hiatt, E. N., Chan, A., Sweeney, M. & Dawe, R. K. Neocentromere-mediated chromosome movement in maize. The Journal of cell biology 139, 831–40 (1997).
9. Longley, A. E. Abnormal Segregation during Megasporogenesis in Maize. Genetics 30, 100–13 (1945).
10. Rhoades, M. M. Preferential Segregation in Maize. Genetics 27, 395–407 (1942).
11. Buckler, E. S. et al. Meiotic drive of chromosomal knobs reshaped the maize genome. Genetics 153, 415–26 (1999).
12. Mitchison, T. J. Polewards microtubule flux in the mitotic spindle: evidence from photoactivation of fluorescence. The Journal of cell biology 109, 637–52 (1989).
13. Gadde, S. & Heald, R. Mechanisms and molecules of the mitotic spindle. Current biology: CB 14, R797–805 (2004).
14. Dawe, R. K., Reed, L. M., Yu, H. G., Muszynski, M. G. & Hiatt, E. N. A maize homolog of mammalian CENPC is a constitutive component of the inner kinetochore. The Plant cell 11, 1227–38 (1999).
15. Yu, H. G. & Dawe, R. K. Functional redundancy in the maize meiotic kinetochore. The Journal of cell biology 151, 131–42 (2000).
16. Dawe, R. K. & Hiatt, E. N. Plant neocentromeres: fast, focused, and driven. Chromosome research: an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology 12, 655–69 (2004).
17. Hiatt, E. N., Kentner, E. K. & Dawe, R. K. Independently regulated neocentromere activity of two classes of tandem repeat arrays. The Plant cell 14, 407–20 (2002).
18. Hirokawa, N. Kinesin and Dynein Superfamily Proteins and the Mechanism of Organelle Transport. Science 279, 519–526 (1998).
19. Sharp, D. J., Rogers, G. C. & Scholey, J. M. Microtubule motors in mitosis. Nature 407, 41–7 (2000).
20. Woehlke, G. et al. Microtubule interaction site of the kinesin motor. Cell 90, 207–16 (1997).

21. Lawrence, C. J. et al. A standardized kinesin nomenclature. The Journal of cell biology 167, 19–22 (2004).
22. Miki, H., Okada, Y. & Hirokawa, N. Analysis of the kinesin superfamily: insights into structure and function. Trends in cell biology 15, 467–76 (2005).
23. Setou, M., Nakagawa, T., Seog, D. H. & Hirokawa, N. Kinesin superfamily motor protein KIF17 and mLin-10 in NMDA receptor-containing vesicle transport. Science (New York, N.Y.) 288, 1796–802 (2000).
24. Bloom, G. S., Wagner, M. C., Pfister, K. K. & Brady, S. T. Native structure and physical properties of bovine brain kinesin and identification of the ATP-binding subunit polypeptide. Biochemistry 27, 3409–16 (1988).
25. Dorner, C., Ullrich, A., Häring, H. U. & Lammers, R. The kinesin-like motor protein KIF1C occurs in intact cells as a dimer and associates with proteins of the 14-3-3 family. The Journal of biological chemistry 274, 33654–60 (1999).
26. Mroczek, R. J., Melo, J. R., Luce, A. C., Hiatt, E. N. & Dawe, R. K. The maize Ab10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. Genetics 174, 145–54 (2006).
27. Kanizay, L. B. The variants of maize chromosome 10 and their roles in meiotic drive. 111 (2011).
28. Dawe, R. K. & Cande, W. Z. Induction of centromeric activity in maize by suppressor of meiotic drive 1. Proceedings of the National Academy of Sciences of the United States of America 93, 8512–7 (1996).
29. Williams-Carrier, R. et al. Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize. The Plant journal: for cell and molecular biology 63, 167–77 (2010).
30. Slotkin, R. K., Freeling, M. & Lisch, D. Mu killer causes the heritable inactivation of the Mutator family of transposable elements in Zea mays. Genetics 165, 781–97 (2003).
31. Martienssen, R. & Baron, A. Coordinate suppression of mutations caused by Robertson's mutator transposons in maize. Genetics 136, 1157–70 (1994).
32. Zhang, X. et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 126, 1189–201 (2006).
33. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews. Genetics 11, 204–20 (2010).
34. Meyer, P. DNA methylation systems and targets in plants. FEBS letters 585, 2008–15 (2011).
35. Gent, J. I. et al. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. Genome research 23, 628–37 (2013).
36. Jacobsen, S. E. & Meyerowitz, E. M. Hypermethylated SUPERMAN epigenetic alleles in arabidopsis. Science (New York, N.Y.) 277, 1100–3 (1997).
37. Bender, J. & Fink, G. R. Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. Cell 83, 725–34 (1995).
38. Stokes, T. L., Kunkel, B. N. & Richards, E. J. Epigenetic variation in Arabidopsis disease resistance. Genes & development 16, 171–82 (2002).
39. Cocciolone, S. M., Chopra, S., Flint-Garcia, S. A., McMullen, M. D. & Peterson, T. Tissue-specific patterns of a maize Myb transcription factor are epigenetically regulated. The Plant journal: for cell and molecular biology 27, 467–78 (2001).

40. Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401, 157–61 (1999).
41. Cronk, Q. C. Plant evolution and development in a post-genomic context. Nature reviews. Genetics 2, 607–19 (2001).
42. Matzke, M. A., Mette, M. F. & Matzke, A. J. Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. Plant molecular biology 43, 401–15 (2000).
43. Jia, Y. et al. Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. PLoS genetics 5, e1000737 (2009).
44. Slotkin, R. K., Freeling, M. & Lisch, D. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. Nature genetics 37, 641–4 (2005).
45. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. Journal of molecular biology 215, 403–10 (1990).
46. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England) 23, 2947–8 (2007).
47. Rozen, S. & Skaletsky, H. Primer3. (1998).
48. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research 18, 821–9 (2008).
49. Koren, S. et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nature biotechnology 30, 693–700 (2012).
50. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics (Oxford, England) 25, 1105–11 (2009).
51. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics (Oxford, England) 28, 1647–9 (2012).
52. Au, K. F., Underwood, J. G., Lee, L. & Wong, W. H. Improving PacBio long read accuracy by short read alignment. PloS one 7, e46679 (2012).
53. Shi, J. & Dawe, R. K. Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. Genetics 173, 1571–83 (2006).
54. Gruntman, E. et al. Kismeth: analyzer of plant methylation states through bisulfite sequencing. BMC bioinformatics 9, 371 (2008).
55. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nature methods 9, 357–9 (2012).
56. Robinson, J. T. et al. Integrative genomics viewer. Nature biotechnology 29, 24–6 (2011).
57. Ambrose, J. C. & Cyr, R. The kinesin ATK5 functions in early spindle assembly in Arabidopsis. The Plant cell 19, 226–36 (2007).
58. Ambrose, J. C., Li, W., Marcus, A., Ma, H. & Cyr, R. A minus-end-directed kinesin with plus-end tracking protein activity is involved in spindle morphogenesis. Molecular biology of the cell 16, 1584–92 (2005).
59. Lawrence, C. J., Morris, N. R., Meagher, R. B. & Dawe, R. K. Dyneins have run their course in plant lineage. Traffic (Copenhagen, Denmark) 2, 362–3 (2001).
60. Wordeman, L. How kinesin motor proteins drive mitotic spindle function: Lessons from molecular assays. Seminars in cell & developmental biology 21, 260–8 (2010).
61. Zhang, Y. & Sperry, A. O. Comparative analysis of two C-terminal kinesin motor proteins: KIFC1 and KIFC5A. Cell motility and the cytoskeleton 58, 213–30 (2004).

62. Matthies, H. J., McDonald, H. B., Goldstein, L. S. & Theurkauf, W. E. Anastral meiotic spindle morphogenesis: role of the non-claret disjunctional kinesin-like protein. The Journal of cell biology 134, 455–64 (1996).
63. Fink, G. et al. The mitotic kinesin-14 Ncd drives directional microtubule-microtubule sliding. Nature cell biology 11, 717–23 (2009).
64. Kalisz, S. & Purugganan, M. D. Epialleles via DNA methylation: consequences for plant evolution. Trends in ecology & evolution 19, 309–14 (2004).
65. Foerster, A. M., Dinh, H. Q., Sedman, L., Wohlrab, B. & Mittelsten Scheid, O. Genetic rearrangements can modify chromatin features at epialleles. PLoS genetics 7, e1002331 (2011).
66. Robertson, D. S. A possible dose-dependent inactivation of mutator (Mu) in maize. MGG Molecular & General Genetics 191, 86–90 (1983).
67. Eichten, S. R. et al. Spreading of heterochromatin is limited to specific families of maize retrotransposons. PLoS genetics 8, e1003127 (2012).
68. Goettel, W. & Messing, J. Epiallele biogenesis in maize. Gene 516, 8–23 (2013).

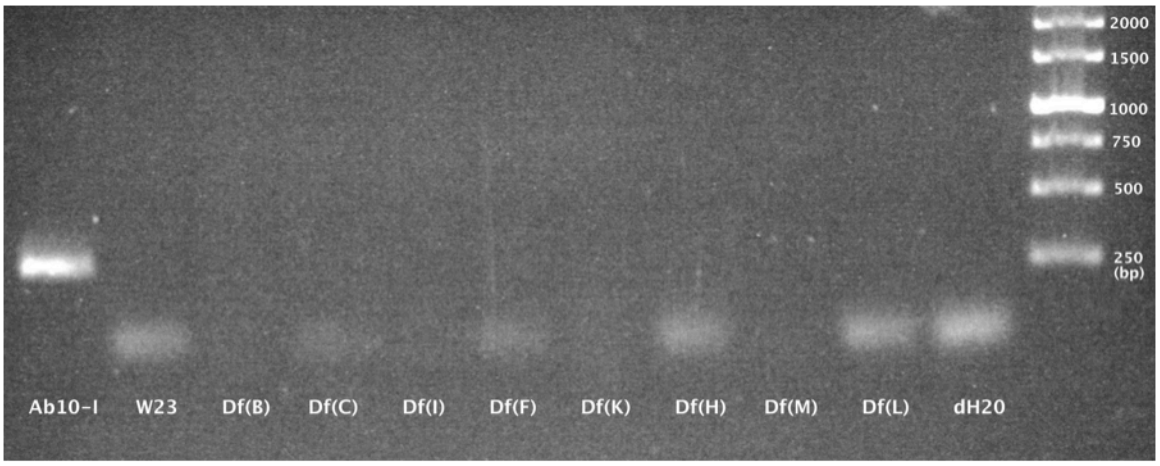

Figure 3.1: Kin618 is not present in any of the Ab10 deficiency lines nor the inbred line W23. *Kin618 (all)* primers used in assay with product size of approximately 200 bp. Lower bands are primer dimers.

>B5contig

TTTGCCGGCGGCCAAGCCTTGTCTGGCAGCTGCTTCTTCCACCAACCCCACGTCATG
TCTAACTCGTAGACTGACGGCCACCATTCAAGCGCCGGGAAAACAGCATCAACAGC
TATGTCGTCATCATCATGACGACCTCTTCCTCTCCTCTGGTCGCAGCAAGGCAGTAA
GATGGAGAGACTCAAGTCCGGCGACTTGCTCTAGCGCCGAAGAGGTCTCCCCTTCGT
TCAAGGATATCCTCCTTGCAGGTTCGGCCTTGGCCCAGGTGTTGGCGGCAACCGGTG
ACCCTGTTCACCAGGACGCCTCGATTGGGACCTTGCAGGCCCCTGTTCTAAGGAAGA
AGGTTGGGACCTTGCAGGCCTGTTATGAGGAAGAAGGCAGCTGAGAGGTCGGACAT
CTCCAAGGGATCAGTACGACATAGCAGCTGGAGCAAGCGTGTTGCGCCCAACTCCC
TTCGCCAGGTGCAGAAGCACCCAAGAGCCCAGTGAAGTTGGTGGATCACGACGGTT
GTACCCAGGTGGGAAGTCGTCATGCGCTGAGGCAGAGGAGGCAACTAGAGCGACGC
CCTCACAAAACATGTCCCGATCGTCTTTAAAGGGTGGTGCTTCATCTGCCTTTCGGCT
CACCATCGCGCCAACGCTTGTAACTCCAAAGTTCGTTGACTCCTTCAGTCCTTTGCCA
TTCTCTTGGACATCGGTCGTTCTCCTGCCCTTGCCCCCTCCCCTCCTCCCTCCCTAGCT
CAGAAAGATCCTGGTGGTCGCGGTCCGCCGGCGACGTCCCTCGGTTGACCTCCGGTG
AGGTGACAACCCAGTCATCCCTCGTTGAGGAAGCAGAGGAGCTGGACGTAGAAGGG
GCTTGTCATTCAAGATGCAGCAGGGCTGACTGCGCAAGTGAAGACAAGGTGGGCGA
TTTTCGGAGACAGGCTGTGATCAGGGAAGGAAAGCTAATGGCTAGTTTCGGCGGAT
TGAGAGTGCCAAGAAGAATCATCAACTTGTCTCCCACGATCAATCAGAGAGCAGAG
GAACTTGCAAGAGCCTTAGTGCTCTCCATTTCTTTGGAGAACCTTCAGGGAATTGAG
GAAGCCATCCTTGCTAATTTCGCCCGAAGGCTGGATGTTGAGCATAATTTGATGTCC
TTGAGGCGCCTCGGTCCCGAGAGTTTCCTCCTTAGGTTGCCTAGTATGGAGGTTGCA
GATCGCGCATATGATGGTGGCATAACAATTCCAATTTCCCCCTCAATTAGGCTCCAT
GTCATGAGGTGGAGCAAATTTCTCACTGCGTTAGCGGCGTCCCTGCCATGCATTGTT
GAAGTGGAATTGAGAGGGGTCCCGGTTCATGCCTGGAACTTGGCAAATGCAGACCT
CCTTCTCAACGAATGGTGTTGGATCCCTGATAATCACCCAGAAGTCACATCCTAGCA
AGAAGCTATCAAGATCGACGCCCGGTGTTCAAGCCCGGATTCCATACCCAAAGAGA
TGGAGATCGAGATCATTGACCCAGTGATTTCCGATGGAGGAGCATTTCGACGTTCCC
TGTTTTACTCAGTCTAGGTGTCAGTGGAAGTCCTTGATCGGACCGAGAAGAATTCAG
CTCCCCGCCGTCATCGCTGTCGTAGGATGATCAGCATGGCCCTAGGAAACGGAGAA
GACGGCATACAAACAGAAGATCGTCCCCGGTCGACGGCACCCTAGGATTCAGGGAT
GTGGCGGACGTGCCGTCGAGGTTACCTATCCATCCGATGGGATGCGGCTCGCCCGCC
TCGACTCCCAATCTACGCGCGGCCAGGCGCGCGCGTCACCCTATATTGGCTCTCGGT
CCATGTGCGGCTGGGCACGCGGCCGCCTCGGCTCCCGGTCCACGAGCGGCTAGGCA
TGCGATCGCCTCGTCTCCCCGTCAAGCGCGGCTGGGCACGCGTCCACCTCGGCTCCC
TATGCACACGTAGTTGGGCACACCGCATCCGCCTGCCTCGACTCCCAATCCCCGCCG
CCCCTGGGGCAGATGCATTGATGGTGAAGACGCGTGTGTGGATTCCTAAGATCCTCA
GCTCGTTTCGAATGCCATTACAAGAGAGGCTGGGCCCACATCTGATATTCCCCAGGC
AACAATGGGGTTTCCGGACATCCAATCTACTTACTAGGAGATCCCTCGAAGCCTAGA
CGCCTGCCAGAATGAATGGGACGCTGAACCAAAGATAGGAGAATTCATTAATGAAA
TCGGCTGAAGCATTCAAAATGAAGAGGTCCAGATGCCCTTCCTCCAGCAACTCTTGA
TTGCAACACTAGGCACCCAGGTCAACTTTATGACAGCATGAACAAGGTTGAATCATT
GTTGAACGATCTCTCGGATATCACATCTTCTCTGATGATGATTGAAATGCATATACC
GCCATCTGTCTCGGAGAGCTCTAAGACTGCACAGTCTGATGGCCTAATTGAGAAAGG
CATTCCAAAACTTGATTCACTAAAGTCACCTCCTTTTAGTTCATCCCAAGCTGCCCCA

CAGCTGATGGTAACTCAAGAAGACGAGTACAGAACGCAAGGTTGGTAGAAGACAA
GGAAAGAGAGGACTCGACCCATCAGATTATTGAGACATTAGTGAAAGGGGTAGGGG
ATGCAGTGGAGATACTGTAATTAGCGCAAGACAAATCCTGGTTAAACCGGCAAATT
GAGTTGATGTTCGAAGACCAGCTCCGCAGGCTGTTGGAATTGCCCTAACACCTGAAC
AGCAGCAGTTCTTCGACCAGCTTACCAAGCAGATTTCTTGCATCCTCCCAACCACGG
CTGATCTGATACTGAAGAATAATAGCAACAAAACACCTGCACCCACTCCCAGGCGC
AGCCGTCGTATTGCAGGGATGGGAGCTGAATTTGATATCCACGACTTGTGCTTAGGA
TCCACAAAAAAAGTCATGAAAACCTTGAAATTGATCAAAGATGGGGAAGGTATCAG
TCAACAAGCCAGAAAGGATTATGAAAAGGTGTTTTCTGAAACACTTTCCCCAGCCGT
CATGTGTGGGAGTGTGTTTCTTTTGTTTTAGGCTTGATGCCTTTTTCTCTCCTATTAAT
ATAATGATACGCAGCTCTCCTGCGTATTCGAGAAAAATAATGTAAATTATATATATA
TTGTATACTATATAAGCAAGGTTTTAAAGTTGTCCGAATAATCGCGATTAGTCGCGA
TTAGTCAGCCTTATCGCCGGAGTAAGCACGATTAGGAGCTTCGACTCGATTAGGCCG
ACTAGGAACCGATTAGTCATCCTAATCATTGACGAATCGTGGTTAGTCACCCGATTA
GGTGTTATGACCGACTAGTTTATGCCTGGTTTGGGCCTGTAATATGTCCTCTGTTCTA
TGGGTCGGCGGCTGGCCCACCATCTCTACAGTAAAAGTAGAAACCCTGTGCTGTGTT
GTCCTACTGTAAACAGCCGCACAATAGCCACCGCCCCTTCTCCAGCGTGCGACTGCG
CCCCCTCTGCTCTGGTGGCCTTCTACTTCGTCGCGGCCCTCTGTTCCATCATGTGGCT
GTTGCAGAGGCCTCTGCTCGAGCCTTTGCCCTCCAGTGTGTGACGCCCGCCTCTGCT
CCATGCGTCCACCGGAGTCCAGCGTGTGACCGTCCTGAACTCCGGAGTTCAGACTTC
AAAGTTCACAGTTCATACTTCCTAACCACTCTAGAGAACATCAGAGTTCTTTAGAGT
TCACGGTTCATAGTCTGTTCTCAACTTCAGACCGAGATGAGCACCTCTTTAGACGAT
AATCCTTTTGTTCTTTTGCTTGTTCAACGGTATTGTGCATGCTTGCTACAACTATAAT
GTACTACTGTACTACAGTCCTGCTACAGCTATATATATACATATGGTCTTGTTATAGG
TGGACGACTAGGAGTCGACTCGGCTCGACTAATCAAGCAAATCGACGACTAATCGT
GATTAGTCACCTTATCAATGCTCAGGCGACTTTAAAACCTTGTATATAAGCATATAT
ATATGTCCATAAACATACGGACGATTAAATGAACGACCAGCTCGGGTCAATTAGTC
GACCTGGTCGACGACTAATCGTGACTAATCGACTAGTCAGTCCCTATATAACTAGGT
TCGACTAGACGACTTAAAAACATTGACATCAACATGTTTATTTCAAATTACTTCAAT
ATCAAACAAAAAAAAACTTAATTCACGCAGAAGTATGCCTTATCCTCTTCATTAATA
TATGTGTCCACTGCACCCTGCTTCTCAGGAAGAGGGGCTTTCTGCTCCACCTGTGGA
CTATACTCCAATGCATCTTGCCTCTTAGGAAGAGGGGCTTTCTCATTATGCTTTTGGT
CTTCCTCAGTTTCCCTGGTCATCCAATCAGCACAAGTTGTTATTTCTATCTTCACAAA
TGTGCACATAAATCGCTTATAAAATAAAAGCTAGCTCTCTTACCGGGCACTAGCTCG
CTGGCCTCCACCAACCCCACGGCCCCCTCTGTCGCGCCCATAACCTCCACGGGAGGG
AGGACCACCATTCGGACACTCGTTACCTGAAACATTTAGAGCATAAGCATTTCGTAC
ACGAAAACACTTTTGTGTCAGAAATGGGAAATGGGTGCCGATAAGAACATGGGCGG
TGGGGAGGTTTGCGCCGCCAGGTTTCTGGAGGAGCAGGGAGCATCCTGGACGGCGG
GGGTGGCCTGGGTAGAGGTGGTCGCGCCCGTGCCATGGTGGAGGAAGCTGGTACCG
TTGGAGACGACGCCAGGAAGGAGTTCGCGCCCTGTCGCGCCCGCGCCGTGCAAGGT
GGTAGCGGCGCCATCCTGGGTGGCACCGAGAAGAGACCCACGGCCGACCGCGCCTG
GGTTGAAGGAGACGGCGCCGACGAGGGGGGGCGGGGCCACAGCCGCTCCGTGCAG
GGTGGCTGTTAGGAAAGGTGAGTGAGTCTCTGGTACGTGGGGCTATGTCCCTGGGTC
ACTAGTCAGTGGACCGTATGTCATAGTATAAGTACGACGATGGTTGTCTTGTGCAGG
GGAAGAAATACTACGAAGAAACTGAATACAGAGAAGCTCTCTGCTGTCTCCTTCCC
AACCCAACCTATCTCTCTCGCCTACCTCCTTCTCCTCTTCTCCAATCCAATCCACTTG

TGATCTGTATGCTAACAATTGGTATCAGAGGAGTCGATTCTCGGTCTAGCTTCATGG
ATCCCGCTGCCCTCAAGTTCCTCCTTGACGAGATCGGCAAAAAGTTCGATGAGGCCG
AGTCTCGACTCGATCGGCGCTTGGCCAGGTTACAGCAGCCCGCACGGGGAGACACG
TCGTCCTCCTGCAGCAACAACTGGGGCGTCGTCCCCTTTCCGATCCAGTCCGCGGCT
GTTAGAGCCGAGGTGGTCACTAACAACTGGGGCGACCTCTTTGACAGTGGAGACGA
CTCCGACGACCAGCGCGACGAGGAGCCCATCATCGCCGACAATTGGGGTGGCCTGT
TCGACCAGCACTTGTGATCTGTATGCTAACAATTGGGGTGGTCTGATGCCAAAGGAA
AGCAACCACGGAGGGGTGCACGAAGAAGAAATGAAAAGCATGTGTTGTCTTCTTCG
TGAGACTCACCGCTAATGCATTTTTTTGGTGCTGCCTAAAGAAATGGAAAGCCTGTG
TTGTACGCTGACGCTGCCCATGGTGCCAGACTCCTTCATGCCCGCTCCCCATCCTTGC
TTGCTCCGGAATGCTCGATATGATGTACGACTTACTGATCTACTACTATTACAATATT
TGCTTTTTTTGTTCCCAGCTGAGCTATGTGAATCATGTGTGCTACCAGCAGAGTTTAG
TTCCTACTCCCACCCTCATCTCTCAGCTTGCTCTGTTCTGTGTCATCTCCTTTTTCTCT
GTTCAATTCAGGAAACCGTTTTTGTTTGGTTGTAATTTGTGCACTGCTTTTCCATTCA
GATAACATGGGCTCCTTACCACTTGCCTCTTTGCTTCTCAGTTCCTACCCTCCATGCA
ACTAAAACTAGGAGCAAATCAATTTTAGTAACTGAAACATTAGAAATCTGTTTTCCC
ATTCAGATAACATGGGCTCCTTACCACTTGCCTCTCTGCTTCTCGGTTCCTGCCCTCC
ATGCAATTGAATCGTTCTAGTACATAGGAACTAAAACAACAAAACGATTTTAGCAA
CCTGAAACATTAGAAATTCGTGTAGATCAAAAGACAATTTGTGGGCATACCAGTTGC
ATTGAATGGATGACTGAAAAAAAATCACCCTCATCTCACATCTCTTAGCTTTCTCTG
CTCTGTTGTGGATGCTATCTTTATATACAGAGACAATATTCTCCGTACATGTGCATCT
GTGATGTTTTTGTACATGTATTCTCTATCGGGGATAAAATGAGAAAACCGAACAAGT
TAGTGTGATCTGATGTAGGCCATAGTGATCGAAGAATCTTAGCTGTTTTACAGATTC
AGATTTCTTATCATAAGGTATCAGAAGTGATAAGGGTCTTATCCTCAACTTCATTTCT
GAAATTATTCACAAGATGAACTTTACACGTAAGTAACAATCGTACAATGAAGAATC
GACACCTCACTTCAAAGATGTTAAACGTTTCCTTATTTTCTGTGCTTAACTTTAGGCT
CGAATGAAACGAACCATTTCCTCTCCTATCTTACAAAGCATATATGTCCACTTAAGC
TTCAACGAGCTTTCATTACATATATAGAAATATGTAGTCTTATAGATGGGTATTTCA
AGTCCTATTTATGCTAATCCAACTCAAAGCAGTTTCATCAAATCATCCAACATTAGT
CCATCCAAAATATTTCCAGATCACCCTGTACCTGTAGAGAACATAGCAGGCTTAAGC
AACTTGTGGATGATAGCCAAATAAGCAATCTATTATTAGTTATTAAAAGGTAACAAG
GTTTAATCCCGTCGGGCAGGATTACTGAAAGAAAACTTTCATGTAACAGATTTATCT
TAATTACCCTTTACGTGATTGACAAAGCTGTGCAAAAAAAAAAAAAAAAAAAAAAA
AATCAACAAAATTTCAAGCACCCTTATTGAATCATTCAGAGAATCAGAGGTGGCAA
AACAGTGGATAACATCCTAAAAGAACCCGCCAAATACTTGTACACAAATGGTTCTTA
CTAATTAGATATTCACGTAGCAAAGTGTCAGACATGCTAATATTGCAGGTGAACTGT
GCATCGTGACTTTCATGAAATCAAAACAGACAGGACCATAAAATTTGTCAGCAACA
CAAAGGCACAGTTCAAGATATGGTATAACAAGAAAAAGCAAAACTAAGCATGGATA
GAAAAAAGGTTGCTGAAGAACTATGATCTGTTTGGCATAACTTATTTCATAGTTTTC
ACAAATTAAGCAGTTGTGCGTAACAACAAAAATCTGCAGCTGCTTAACCAAACTAG
GAACTAACTGAAAAAGCAAACTACACTTTTCACCAGCTTCTCATAAAAGAAAATTGT
GCTCAATTACAAGTTGTGCCACACAAGGCCTATACTACAGTTCAACTCTCTTTACAT
ATCTAAGAATCAAAAGTACTGTTACTCTTTGCTGTTACGAATCAATACGACACAAAG
TAGATCAAACAAAGAGAAGACACAGATTTAACGTGGAAAACCCCTCCAGGCGAAGG
AGGAAACACCACGGGCGCCGGCCCAAGCCTCCCAACGACGGGCCTCCGCTTCGCTC
GCCAACTTGGCCGCTTTCTTCAGAATTTGAATCACAAACTCAACACTTGCCTCTAGC

AAATAAGTTGTTTGGTACAAGAAATTTTTTAATTATAATCAACATGAGGTCCGGGGA
AACGGATAAACCGAGGCAAGCAATCCTCCCGCGAATGCGGGTTTGAAACGGCCTTT
CCGCATTTGAAGGGGAAAGCTTGACTCAGTTTCTCTCTTTCCCAGACTCTACTTATGT
GAGAGTCTTCGGCACTAGGTCTATCGTTTTTTATTTACCATAGGACAAAACTTGAAA
GATCTAACCACAAGATTTCAACATTTCTTACAAGAATATTTCTTTGCAGTATAGCAA
GGTTCATGCATAACAAATATATTAGGGGCACTTGCAAAATTGCAATTGATTCTTAAA
AACCATGAAAATAAATAATATTGTAATAACAGGACAATGTTTTAGTTATAGCTGAAT
GTACCATGGAATCTCAAGGAGCAAGACTCTTTTTGAGCTTGAGGGTTTTAACATGTT
CACAAAACCACCGATGTTTCTCTTGCACTTCAGCTATCGCCTTTTCAGCAGCCTCTTC
ATTCTTTGCTAGAAGGTATTTCTCGATCTGAACAAACCACTCCCCCAGTGTCATATTC
TCCAATCTGTGTCTTCCTTTGGGCTCCCCGACCCCAACATGGTCTCTCTCTCCACTGG
CACCAACAGCTTCCCTCTTTTCATTCTTTTCCACTTCTCCAACACCATCAGGCTTTGA
CCCTGGTGAAACCTTTTCCCTCAATTCCTGATATACTTCAGTTTCCTCTCCCTTGCTCA
AATTCTTCACCTCCTCTTCTTCCACACCACCATTGTCAGCAGGAGCATGCTTAACATT
GACACTCCCCTTTATTGCCCTCTCTGTCAATGTCAAATCCTCCATCTTCTCATCCTCC
ATGGATGCCTTTGAAGTTCCATCCAACTTGTCGTCCTCAGTTGCCACCTTATCCACTT
CAGAACCATCATTCACGGCCAACACCCTCAATGGACTCCCCCCTCGTTCTCTGCCAT
CAGCCATATCACTTTCCTCTTGCTCATTTTCATCGTTCTCAACAACCGCATCCTCAGC
CTTCTTTGTACTTGCCCTCGTAGCCCTTCCTCGGCCTCTGCCTCCTCTGCTCCTCTTTG
CTTGTCCTCTCCCAGCCGCCGGCACCACCTCATCCTTGCCCCCGGCAGCTGTCCCTTC
CCGATCTCTGGCCACAGCCACCTCACCCTCATCTTCCTCCTGAGTAACACCTGTATCA
TTCCTTGCACTTGACCTTGTAACCCGTCCTCGTCCCCTTCCGGTCCTCATCACTTTCCC
CTCATCGTTTTCGCACCCCAATACGTCACATTTCCATGCAGCAGCTCTCGTTGACCTC
GTATTTTGCGGCTGTGGAGGAAGGACGGCCGTGGCTGCAGCCTTCCTCGACCTGCCA
CGGCGCGGTATTGCCGCTTCCTCCTCTTCATCTTCCGCTTCCATTTCCACCCTAGCGG
GCTCAACGGTCTTCCTCCGCCCGCCTCGGCGAGTCGGCAATGCAGCTTCCTCCTCCA
CTTCATTCCCAGCTTCGGGGGCTTCCGTCGCACCAGCCTTCCTCCGTTCAGCCCTGCG
AGGAACAGCCGGGCCCTGTTCATCCGCCGCCGCCGCTGCGTTTCGCCTGGAACGCCT
TGTCGCGACGGGGCCCGGTCCCGGGCCCGGGCCTGCATCGGTCGATATGGAGACGG
CGAGCACGGTTGTCTCCCCGATCTTGATGAGGTCCCCGTCGGAGAGCGGAGCGGGG
ATAGTGGCCACGAGAGGCGTATCGTTGAGGAGGGTGCCGTTCGAGGAGCCTAGATC
GGTGACGGTCCACCGGGCCGCAGGCGGCGGGAGGAACTCGACGGAGAGATGCCTCT
GCGACGCGCCCGCGTCGCCCACGGAGAGGTCGTTGCCCTTGGCAACGCGGCCGACA
CGGAGCGCCGCGGTGCCGGCGCAGCGCTGCAGGGTATTCCCCTTGCAGGGGCCCTT
CTTCACCAGGAGAGTAAGGACAGGCGGCGGATCCGCCATCGAGGGGGCTCAGGGGT
CGGAGGTCTGTGAAGAGGGAAGCGAAAAGGATTTCAAATTTCTGGGCAGAGTTGGG
GAAGAGGACTGAACGCTGCTGTGGAGTGTGGGAGACAATTGCCATTATGAAACACG
AGAGACACGGTTTCGACGAAATGGAGGCGTTGGGGATGGCTTCTGTGTTTTGGAGA
CCCACCCGTCAGACCTACTCCAAAATGGACACGAGAGAGTCATTTAGGACTGCGCA
AACTTAAACGAGCGTACCTTTCACCATTACCGTCGCCATGCTCAGACGCCGTCTGTC
GTCGTCGTCTTCTCCCCTTCGCGGCGGCTGAGCGGTTTGCATAGTGGTCGCGGCTGA
GCGGCTAGGAGAGTTCTCGTGTCTTGGAGATCTCGCGGCGGTGAGCTCTATTCGCCC
TTCTTACCAATTCAACTCTCTCTCTCCCTTGTGTTCTTCCTCTGACTGTGACCGGATTT
GCGCTTAGGCTTTCTCTGAGTTCGGTTTCATGGATTTAGATCCACACGAAATCTCTGC
AGTCTGCCTCACTCGTATCAACTCTCCCGATCTACAACTAGAGCCTGTGCCATCACC
GTTTACCAATCTGGCGTCACCTAGTTCGTCTGCGCCAGTGGGGAAAATAAGTCGCAG

TGAGGTGAGATCAATTACAGAACTATGTCATTAAATGTGAAATTAATCTCGCTGATT
CTTGGATGCTCCTGCATTTTCTGTAATATAGGATGTTCAGAGGATTGATGCAACATC
ACAACTGGACTGGGAATCGATTGTAGTTTCTGATGTATTGGATGAAGAAGGAAGAG
TACAATTACCATGCGAAAATGAGATATATTGTAATCTTGGACTCAATAAAGAGGATG
AGGCTGCGAATAATAGAATGCGTGATAGTGTGGATATTGGGAACGAAGACCACCAT
GCTGATCAGCCATGTGAAGACTTCATTCCAGATGAGAAGAGGGTGGAGTACAATAG
GATGAATCCTTCAATGCAGCCAGGTTGTTTGTTTCCTAACATGAAAGAATTTAGGAT
TGCTATGCGTCAGTATGCAATAAAACATGAGTTCGAGCTGGGAATTGATGTTACTTC
GACAACCAGATACGTTGGATACTGTAAGGGTGGTGATTGCCCTTGGAGGATATATGC
ACGCGAAGAGAAGAAAGGATTGCCTACTATTGTTGTAGCTGTACTAGACGATGTGC
ACACTTGCACATCAAGTGGAAGGAGGAAGACTACTACGCCAACTAGTGCTTGGGTT
GCATTCCACGCTAAACCCCTGCTCATGAATAAACCACAAATGGGTGCTAAAGAGTTA
CAACAAACACTGCAGGGAACTCATAATGTGACTATTGGGTACGATACAGTTTGGAA
GGGGAAAGAGAAAGCCTTGAAAGAGTTGTATGGATCTTGGGAGGAAAGCTTCCAGC
TCTTGTTCGCTTGGAGGGAGGCTGTACTTGCGAAGATGCCCGACAGTGTGGTTGAGA
TTGATGTTATTTTGGAAGATGGGAAGTACTATTTTAGTCGATTCTTTTGTGCCTTAGG
TCCATGCATATCAGGGTTTCGAGATGGGTGCAGACCTTACCTTAGTGTCGACTCGAC
GGCATTGAACGGTAGATGGAACGGGCATCTAGCATCAGCTACTGGTGTGGATGGCC
ACAATTGGATGTACCCTGTAGCTTTTGGTTTTTTCCAATCAGAGACGGTTGAAAATT
GGAATTGGTTCATGACACAGATGAAAAAGGTCGTGGGTGATATGACAGTTCTTGCTA
TATGTTCAGATGCACAAAAAGGCCCGATGCATGCTGTAAACGATGTTTTCCCGTATG
CTGAGAGAAGAGAATGCTTCAGACACTTAATGGGAAATTATGTCAAACACCATGCT
GGTTCAGAGCACATGTTTCCAGCAGCGAGGGCATATAGGAGGGACGTATTTGAACA
CCATGTCAGAAAGGTCAGAAATGTTCATAAGATTGCTGAGTACTTAGACCAACACC
ACAAATTCCTTTGGTACAGGAGTGGTTTCAACAAAGATATCAAATGTGATTACATCA
CAAATAACATGGCTGAGGTTTTTAATAACTGGGTTAAAGACCACAAAGACCTTCCTG
TGTGTGATTTAGCTGAAAAAATTAGGGAGATGACGATGGAGCTGTTCCATCGGAGG
CGTAGAATTGCTCAGAAGCTTGAAGGAAGGATTTTGCCATCTGTCTTAGCGTTGCTA
AAGGCAAGGACTAGAGGTTTGGGTCACTTGTCCATTGTAAAATGTGACAACTACATG
GGTGAGGTACGAGACAACACAAATTGTATGACAAAACATGTGGTGAAAGCAGACTT
GAAACAGTGTTCTTGTGAGGAATGGCAGCACACTGGGAAACCATGTCAACATGGTT
TAGCCCTGATAATTGCCCAAGATTGCAGAGATGTAGGCATGGAAAATTTTGTTGACG
ATTATTATTCTGTTCAAAGATTCAGGATAGCATATTCAAGAAGGGTGGAACCAATTG
GTGATCGATCATTATGGCCATCTGTTGATTTCGCCAGTGGAGTGTTTGCACCTATTGC
TAAAAGAGGTCTTGGTAGGCAAAGAAAAAATAGAATCAAAAGCTGCCTCGAGGGTG
GGAGTGGTAGAAAAAGTAGTTGCAACGACAATGAAAAAACGAGAAAGCGAGTTAA
AAGGCAATACACTTGTCCTAATTGTGGTGAATTAGGACACCGCCAATCTAGCTACAA
GTGCCCTTTCAATGGCACAAAAAAAGGCTAGTTCTGAACTTACTTATATATGCATAA
TTTATCTTAGTACTCATGAACTAAATCTTTGAATATTTTTTTGAGTAGGACAAGGAAA
CCACGGAAGAACACCACCAAAAATTGGATTCCTAAAGAGTTGCGTACTTCACAGAC
CGTACCTGAAGAGCCAGAGCAAGAAGATGTTGAAACGCCAGATAGTAACCATGTTG
TTCCATGACCGATATCTGGACAGTCACCATGTTGTTAGTTTCTGTTAGTTTCTGGACA
GTAGCCATTTTTTGTTAGTTTCTGTTAGTTTCTGGACAGTAGCCATTTTTCTGGACAG
TAGCCATGCTGATCTACACAGTAGCCATTTTTTGTTCAGTTTCATGACCGTGATCTGG
ACAGTAGCCAAGACAAAAATGGACAGTAGCCATAACTGAATTCACCACACATAAAT
TGGAACTGCTTCATAGAACAGACATAGTAGCATACAAAATACCTTCATCCAACAAAT

GCCTACTGCTCCACACAGATGACACTAATGTTTCATGACAACAAACAGAAACTAAG
GTTCTACATAGTACATTTGGAGCATTACATACAAAATGGAGGCATCCAAACCAAAC
ACGAGCTCACAACTACATGTTACGATCATTGATCTCTTCGAACCTGTAGTTTATGAC
ACGATAGTGCTCTTCAAATGAAACATTAGCTTTGGCCTCTTCCCACTTCAGGTCATCG
AAAACCATTTCCCATAGCTCAGGATCTTCTGTACTGTAGCCATCACCAATTTCATCAT
CTTCTAAAGGATGTTCATCATCTACGGAAGGGTTGTCATAAAGGTCCCAAGTAGGAT
TCTTTGTTTCCAAATCTTGTTCACAGACAATTGAAGTTTCCTCTAGCATCTTGCACAC
ATGTTCCAGATAGCCGCAAACCTTTCCTATCTTGGCACCTTCAACTTCTGGGATAGTG
GAACACAGTTTTTTAACAGATTGCAACATGGTACATTGCCTCCTCATTTCTTGAATGG
AGTATCGCAACGAATTCTTAACGTTCACATGATACTTGAAATCGACATTTTCCTGTG
CAACAACCATATGCTAATCAACCATATACTTGAAATCGACATAGTTGGCGTAGTAAC
ATAGCAATAATAACACGGCGTACCTTTCACAATAGTAACATAGCAATAATACGAAC
TGTATGAGTTTCATAATACGTACCACATTTGACGAGACCTTAACTGCTTCGACCTTGT
CGACGGCAATGGCGGCGAACTCAACCGCAGCCTTGGCCTCCTTCTCTGCAATAACTG
CCGCGCATCTCGCCTCCGAAGCCGCGACGGCAGCAGCACGCGCCGCTTCAACAGCA
TCAGCAACCAAATGGCTCGGATGAACATCCATCGCAAGAGGAGCGTAGATGGCGGC
TAGGTTTCTCGTTTCTGGAATGGGGAATACAAAGCAAGAAGCGCAAGCTGAGATCT
CGCGCAAGCTACGGTGGGGAATACAAAGCAAGAAGCGCAAGCTAGGCAAGACCGC
ATGAGCACAAAGTCGGAAGGGGGAATACGCGCAAGCTACGGCTAGGGTTTCTGGAA
TGGGGAATACGCGCAAGCAACAGCTAGGGTATCTGGATTTGGGATGGCTTCTGTGA
AGCAAGCAACGGCGGCGCCAGGATTTTAACGGCGTGAGAACAAAGAGAATTAAAC
GGTGACCACAGACGCAAATGACTGTCTGATGTCCATATCAGCAAAAGTCTGACGGG
TGGGTCTCCAAAACACAGAAGCCATCCCCAACGCCTCCATTTCGTCGAAACCGTGTC
TCTCGTGTTTCATAATGGCAATTGTCTACTCCTGGAGTCGAGCCGTCGAGCCGTCGA
GGACTAAACCTATATTTTACATTTCTACTAACGTGGGCGTCTTAGGCTCTTTCGACCT
CCGTTGGCCATTCCAATTCCATCCGAATCGAGAGGCGCCAAACCCCCGCTAACCCCT
ACTCCTAAGAACCCGGCGGACGCGCAGCCCGGAGAGCGCCAGCGCACCGCTGCCGG
GACCGGGAGGCAGCCACTCGCGGCCGCGCCACCGCCGCCGGAGGTCCCCCTTTCCA
GTTCCATCCGAATCGAGAGGCGGAGGCGCCCGGATCCACCGCTCACCCCCGCTCCTC
AAACCTCCATGGATCCACCCGCCATGCCCGCGCGACGGAGGCGCCCGGAGGTTCTG
CACGACAAGGAGAACCCGGCGGACGCGCAGCCCAGAGAGCGCCAGCGCACCGCTG
CCGGGACCGGGAGGCAGCCACTCGGGGCCGCGCCGACGCCGCCTGAGGCCCCCCTT
TCCAGTTCCATCCGAATCGAGAGGCCGAGGCGCCCGGATCCACCGCTCACCCCCGCT
CCTCAAACCTCCATGGATCCACCGCTCACCTCCGCTCCTCAAACCTCCATGGATTCA
CCCGCCATGCCCGCGCGACGGAGGCGCCCGGAGGTTCTACACGACAAGGAGAACCC
AGCGGACGCGCAGCCCGGAGAGCGCCAGCGCACCGCTGCCGGAACCGGGAGGCAG
CCACTTGCGGGCGCGCTGCCGCCGCCGCCGCCGGAGGTCTGTCGCTCCATCCCTCTC
TCCCCGTCCTGGATGAAGTGCTTCAGTTAGGTAGTGTTTGCTTCCAGATTCTTCTGGT
AGGATAGAGCCGCTCCGTTCTTCTAGTAATGTTAGTAACTAGTCACTGGCGCTCCAG
CGCGCGCATATAATCATTGTAATATTAATAGAAATAACACTAGAGAAATACAAACA
CTGTGTTGTCACATTGTTACATCTATGTAGGTTGCTTATCAAAGCATAATAACAAAA
GGTGGTTCTCACATATTTACATCTAAGTTGGTCGCACATGACTGCATAGTAGCAAAA
GGTTCGTAGTTGATGGTATGTTATTTATTGAATGGAGCGACCATGGAGTATTGGTAG
CATCCCAGTCATCCAAAATACGATATTACTTTACTGATATGTAATGAAAAATGTGTT
TCGGTACCTATTGGCAAATTGTAAGCATGATAATGTGATTTGTATTTGAGCAGTTTTG
TTGGTAGTAGTAACGCATTAGAAAAAAAATTATAGTTGTTTAGTGAGGAGGGACAA

TAAGATATATAGTTATGTAATAAAAGGAGATCTGGTTTGCAAAAAACAATCAATAA
CTGGTGTATACCCGTTATGAGTTTCATATTGCTTTGGTACTGCATGTTGTCATATTTC
AATGGTGAGCACAGGTAGAAACAAAGGTGAAGTTATAGTTCAAAGCATCAATATGA
CCTTGGTTTATTCTTTTTTCTGCTTAATATATAATGGGGCGCAGTTCTCCTGCGCGTTT
GAGAAAAAAAAAATAAGCCAACTATTAAGCAATCATATGTGATACATAAGACATGT
TTTAGCGAATAACCAGGTGGGCGCTATTAGGAAAAAATAGTAAAAATGGAGCACAC
TGTTATGAAAATGAATGTGTAATGAGTTGTATCTTGGTCAAGTAGCTTTTTTCCCCTG
TGCAGATATGTATAAGTTGACAAAGCTCATGTTTCGAAACCAAATATTAGGATTTAC
TGTCAATTTGAAAATACCAAAACAGTAAGCAGTCAATGCACGGACGTTTCTGTGCGA
AACTACTAAGGTACTATAAATACTAAATGAATGTTGTGTTTAGAGGTAATGTAGCAA
ACCTGTGAAATGTTTTTCAATAAATGTATCAAGTAATACATATAGATATTAAATGAA
CGTTGTGTTTAGAGGTAATGTACCAAACTTGTGAAATGTTTTGAGTAAAAGTACCAA
GCAATAGATAGTTATGCAAAGAAGAAGTACATACAGAAATAGTTATGCAAATAATA
AATACGTAATAGTACCACCACTGAAGAGAAATTGTTAAATTGGAAGGCACTGTTAG
GGATAGGATTGCGTAGTGAAATGTACCTTAATCAAGCAGCCTTTTTTCCCTTGAGTA
AATATGTTGAAGTTGATGAAGCATGTGTTTCCAAAACTAAGATTAGAATTTACTGGC
AAATTAAAAAGAAGGAAACCATAAGAAGTGAATGCATGGATAATTCTATGATGAGA
TACTATGAACATCATGTATACAGAAAATGTAGCTACCACTTGTATTGTTTAGAGTAA
TTGTAGCAAGCATATGGTCTATACGCAGCACGGCACAACAAAATCAGTATCACTAC
ACACCAAAAGAATGCTTTCTGACGAAATGTAACTAAAGTTCAAGAGGCTGTGGCTG
AGGATGGTATAGCAAACTTAAGCGAGGACACTGACATGTACGGAGCTCTGAGGGTG
AGGATGTTATAGCTAACTTATTTTTGACCGGACGCCATTAACAATCTGGAAGCACAT
GAAGGTCCAACGCGAAATGAACGGACGACTGATAGCTGCGAGTATGCGGCGAAGCA
CAGGTAGGGATGTCAGCAGGCGTCGGCCCGGATGGTCCCTGTCGGCGACGGCGCGG
TGAGTATGCGGCGAAGCAGGGATGATGCTGCGGCGGACGACACCGGTGTAGTGCCA
AGGGCAGTGGACGGTGGAGGAGGGGGGACGTCTAGAGCGGCGACCGACGTCATCTC
CGTGGAGAAGATGGCGGCACGACGGAGACGAAGAAGCCCCTGACTTCTTTCCCATT
CTCTTCGATCGATGATGTAGAAGTTTAATTAATTCGTATAGCATATGTACGTTCGTAT
ATACAATTAATATAGTAAAATAACATTAGTATGCTTTCTCAGCTTGTTATATAATTCA
ATACCATAGGGATCTGCTACACTGTAACCAAACACAGGATGGAGCAGCTCCATTCTA
CTTCACTATAGAACCAAACTGTTCTAGTTACCAAACATAGAACATAACCGTTTTATT
TAGAATCTAGAATAGAACCGTTCCATTCAACTAGGCTCTAGAACCAAACGCTACCTT
AGAACCTTATACCGTCGCCTACGTATCTCCTATTACCATCCTCGTGGATGCGTCGATT
CAAACTCGGCAGAATGCATAGATGCCATTTTTTGGAAGGAGAAAAAAAAGACAAAT
AGTAGCTCACAACTCACGGTTTCTCTTCATGAGTTCCTCTTCATGAGCTCATGTGAAC
TGGGCGTGTTAATTTCTGGCATATGCTCATTCTCTAGTTCAACTCTGGACACTAGGTG
ATAGATGCCTGTTAATGAATAATGCAATCCTTATTTGGTGTTGTTTAATTGAAATATC
CTGATGTATTGATTATGCTGCTGTCCTTCTAGTCTTTGCAAAATGAAGTTGATTTATT
GAAGGAGGTTGAAAGTCTTAAAATCGAGCTTCAACATGTAAGGCAGGACCGTGACA
TGTTACTGTCTTATATGGGAACATATAAAGAAATGACTGAAAAATCTGTCATTGAAT
TAGAAAACGCCATGAAAAAGCCTCTACCCTTCAGGTATGACTATGTGTCTAAATAAT
TTGATCGTTATCTAGATAAATGAAACTGGCTAACAAGTTTTCAATGACATAGGAAAC
ATGTTCGTCCCAGAGAAATATTATAAGAACATTGGAAAGTAAGCTTGCAACTGTGG
AGAAGTTAAAGGTGCGTACATACCTATTTTTTTCATTAATCATTACATGTAATGAACT
AATGGTGTTGTGTTATATCTCAGTAAACACAAGTTTATTTGTATGTTTCTTAACGTAT
GAAGTGTTCACATTGTGTAGGAAAAATATTGCATAGATAATTAATATCTCGAACCTT

CTCTTAAATGTACAGATGTCTAATATGATGTCCTCAGAAACCATGACTGAGTATGAG
GACATGAAGAAATTGCTAGAGAGTGTCCAGTTACGCCTTAAAGAGGCTGAGCAAAC
AATTTTGAATGGAGAGATTTTACGCAAAAAGTTACATAACACAATTCTTGTAAGAAT
TCTTTTAAACTCAAAAATCTAGTATATGTACACATATGCTATCCGGCTATTCCTGAAT
TTGTTTTTTTTTTGCAGGAGTTGAAGGGAAACATCAGAGTATTTTGTAGAGTAAGGC
CTTTGTTGCCAAATGAATCTGGAGCGGTCTCATATCCAAGGAATATACAAGACATAG
GCCGAGGCATTGAGTTGTTACATAATGGTAGAGTCATCTTTACAATTATTTTCGTAGT
TTTAGATTTTGATATGTTGTATTTTAGAATATTGATTATGATATATTTGCATTTCTTAT
TTTTTTAGCACAAGGATATTCATTTACTTTTGATAAAGTATTTGATCATTCCGCAACG
CAGGAACACGTATTCACAGAAATATCTCAACTCGTGCAAAGCGCTCTAGATGGCTAT
AAGGTGCAACATTAGGATGACTATAACTTAGATTATCTTTGCAAACTCTAGAAATAA
TGATTTGTTGGTTTTTTTTTCCAATATCTAGTTTTCTTAGAATGCAACTAAAAAAATG
GTAAATCATTTTAGGTCTGTATATTTGCCTATGGGCAAACTGGTTCTGGTAAAACAT
ACACAATGATGGGCAACCCTGAGTTAGAAGAACAAAAAGGAATGATACCACGATCT
CTGGAGCAAATTTTTCAAGCTACCCAGGCACTTAACTCACAGGGATGGATATATAAG
ATACAGGTGCTTTTTTTTTTGGTAATTTACTTATTTAATTGAACTGACCAGTGAACAGG
TTTTTTGTAACTGTTATCATGGTGTATTTTAAACTTTCAGGCCTCAATGTTAGAAATT
TATCTGGAAACCATACGTGACTTGCTTGCGACGAAACACATTTCTCAAGATGGTGGT
TCTTCAAAGCACAATATCAAGCATGACGCACAGGGTAGTATTGTGTCGGGCCTTACG
GTCATTGACGTAACAAGTATAAGCGAAGTTTCTTCTCTCCTTAGGCGGGCTGCTCAG
AGCAGGTTACACGCTTATTTATTTTCAAAGTTTCAGCATTAAGTATTGGATCGCAGG
AGTATATAATAGCAAAATCTTGTACTGATCACGAATGTTTATGTATGTGCCATAATT
TTTGGCAGCAGATGTAGTATATAATGTAGCTATGTAAAGAAGTCACTACATCATATT
CCCTTGACTGTCATATCACTGCCACTGCGTTCTCAAACTTCTTTTGGAGTTCTAGATT
TTGAGAATTGAGTGGTGTAGGACAGTCATCTATCAATTTTGTTTCCAGATCAGTTGG
GAGAACACAAATGAATGAGGAATCATCCAGAAGTCACTGTGTTTTAACTCTACGGA
TTTTTGGTGTAAACGAGGTATATTCTAATTCACATGTATCGTTGTTTGGCATGTTGTC
TTTTCGTCTATAACTGACCGTTTTTTCTCTTCAGAGAATGGACCAACAAGTGCAAGG
AGTGCTCAATCTTATAGATCTAGCAGGCAGTGAGCGCCTGAACAAAAGTGGTGCCA
CTGGTGACCAGTTGAAAGAGGCTGTGGTATTTTGTTATCTAGTCCCTTAGCATTGTTT
GTGTTTTAGCTTGATATATATAAATAAAATAACGAAGAGGCATGGTTTGTTCCTACT
TAGGCCATTAATAAAAGCTTATCATGTTTAAGTGATGTCATCTTCTCAATCGCAAAA
AAAGATGAACACGTTCCGTTTCGGAACTCGAAGCTGACCTATCTGCTCCAGGTATTC
CCTTATCCTCTTGCATGCCATCTTGTTTTCGTGAAAGTCAAATTGTAAATGTATCAAC
TGATCTCCGCTTCCCAACACAGCCATGTCTCGGAGGGGACTCCAAGACGCTGATGTT
TGTGAACCTATCCCCTGAAGTGTCTTCGACTGGGGAGTCTCTCTGCTCATTGCGGTTT
GCTGCGAGGGTGAACTCGTGCGAGATCGGGATCCCTCGGCGCCAAACCAGAATGTC
GTCTGACCGACCAAGTTGAAATAGGACATCAAGTACTCAGTAAGTCATCATTGAAC
ACCAAGTCACCAACGCAGGAGCTATTAGAATCATGTAATGAATTCAAGTATGTTAAT
AATTATCTCACATTGACTGCATCCACATATTAACCGATATCTGGACTAATACAACAA
ATCTGGAATATCATTATATTTCTTTGAGTGTTATTCTGACTGGGTCAGGTGGCGCCCT
TTCTGGTGCCCTATCCAACACAAGGGGAAAACAGTAATGCAGAGGATCTGAGAAAT
TATTCAGCACAATTGATGTTTTAAAGGTTGAATGCCCCACAAAATTGGTTAGCGGCG
GGAGACAGCTAGAGCAACAAATAAAACATAGTAACAAACACAGAGAGAGCGAGAC
TGATCATAATTCGCATTAAAGTTGCAACTGTTTCTTGTTTCAACACAAGCCACATAC
ATAGAGGTGTAGCATAGGCCCTCTACAACAGGGAACCGACCAAAATAAGGTGACAA

GGTCCGTACGTAGGGCCAAACACCACAACAGGAGGTATGTACATGATGCTGGTAGG
GGTGGTAATGGATCACGATCCAAATGTTTGTCCACAATTTGTTAGGGTCCTAAATTA
ATTTTAGTTCAAAAATGAATAGAAATAGAGCCTGATCCTAACCCGATCTGATCCTTA
AATTTTAGCGTAAAGTTTAGATTCCATTGTCAACCCTAGGTACTGTAAGACATACTC
TAGCTAATGCAGTCCACATAGTAATGTACATAAACTTTACCAAACCACAAGAATGTG
AGTATAAAAGTGCCATAGATATTCCCCGTGATGCGCTTTTCTTCTCCAACCCTGCCTA
ATTACTTCTAGTTAAAAGATACACGGTTTTGCTGGCTTCTGAGCGATTCACGCGGAT
TCGGCGTGTTGCGCGGCGACGTGTGCCGTGGTGGGGTCGTGTTTGGCGGGCTTCGCG
CCATCCGTTGCGCGACGCGCGTCCTCCGTGGGTCCGACGCACGCGGATGTGAGACCT
CGCTTTCGACGGACGCGAAATTTCCAGCCACACACCTTCTCTCGCAGTACAGACGTA
CAGTGCAGGCACGGTCGCGGCTGTTCTTAGCTGATGGCGAACTGGCGAAGGCCACG
ATGCGGGACCCGTCTCTGCTCCATGCCTCCATTACTTCGTGCCATCCTCCCTCCTTCC
ACGCTCCTTCCCGCAAGAGCAGCTTCCCTCCTCTCCTTCCCCCAAACGCTCACCTCCT
CTTCTCTCGCGCTGAACTCGACCTAGCGTTCGACTCCCTCGCGGCCCCATCGCAGGG
TGTCGTCGTCCCTGCCACAGTCCGTCGGGGTCGGCTTGTCCTCGGATCTCGCACGAT
TTAGGAGCAGCGACATCAGATGTTACGGACGCGAGGCCATCAACTCCGCCGCCACC
AGCGCCGATGCCGGCCCCGGCCCGGCGGATCCCTCCCCGGTGGCCCCGGAGGCTAC
AACCTGGCAGATGTCGGAGGTGTGGAAGGAGATCTTCGCCTCCGTGCCCGTCCAGTT
CCGCTTTTCGCTTCTTCAACAGGGAAAATTTCCTTTCTATTGTCACCCCCTGATCGTC
CAGTTTCGCTTAAGTTTCCTTGCTATTGTTAGCAGGAGGCTGTGGCCCGACATACGCT
AGAGGATTTCCTGCGCCAGGGGGGTGGTGATGGAGGGGAACATGGCACTGGCCTGT
CACATATTTGCCACCTGCAAATGGCTGATTCTGATTTAATCACAAGGAAATAATATG
GGTTGTCATTATTTGGATTATATCTTAATGTCCTGTACATGCTGGATCAGGTTGACCT
TGATCCCTATGGGTCACTAGCTGCATTCCTTGATTCCATGGTTCAATGTGTAGCAGAT
GGGGGTACGTTAATGTGCTCAGCTATAGACATGATAGTGCTTGCTGGTGGAAATGCA
GAAGTTTGTTTTTCTAAATGAGAGTTTCAGTTTTAGGATTAAATACGAATTGTTATCT
GCTACACATTGTTTTCCTAAAACCTTTTGCTTCCAGCGTCTGCCTCCACAATACCAAC
TTTTGCGAAACACCACTCCTACTCTCCTCAATAAGTACCACATCATCGGCAAAAAGC
ATACACCACGGTATATCTCCTTGTATGTCCATGTAAAAATTTAGTGTCTTCAAAGTTG
TCCTTTACTGATTGTTTGGACAAGAAAAATGCCAGTATTTTTTAGGAGAATTCCAAA
CTATGATCTACGAAAACAAAATAAAAAGGTCTCCAGTTTATGTTCATAATTTAACCT
TTTTCTAATATATGAAATATGAAACACAACTAAATGCCATATGCAGTAAGTTCCTCC
TGACATCATGTATTATTTGAGCTATAACTACTTTGTTCTTGACGTAGATAAAAGAGTT
AAGTTGCTTGCAATAAACTCACAACCTGTGTTATTTCAAGTTACAAAAAAACCATGT
TACAGGTGAAATATGTATCCAATTCAAAAGTATCATAAATAGCAAATCCAGCAAAG
ATCCTGTGTAAACTGGACTGTCCTTGCAAGCTCTTGATAGCCACATAAGTGTGAACC
AGATAACAAACTATTATTCTACATGCAGTTCAATAGGAAGCATGGTCACCTTTCGTA
AAGAAGTAAAGACTATAGTTCATCCCGGTGAGGTCTCAGTTTTAGGTTCCCTTCGAA
GTATATCACCTGCTATTGGGTTAACCTAGGGTGCATGCAGACAACCAGTTGCACTGT
GGGATCTAATCATTTGTGTTTATGGTGCCGATTTTGGTTCCTCATTCTCGCTTGGTTT
ATCTGGGAGGTCTGCTCCCTGAGAGAAAGAACAAAAACAAAAGGAACCTAAAAGA
ATTTCTGTTATTACATCCACCGTATGGGTTGTAATATACTATGTCTGTTTTTTTGGTTC
ATTGGTTTTGAAATTCTACTATGATATCTATACATAGTAAGTGAAATGTTGAGCCTTC
AGGACATCGGTTGCTACCACCTGAGGCTTTAAACTAATCTCAATCACTTCTGCTTTTG
CATATTGGTTTGGGTGCATGACATAATTGGTCAAAGCAATGCTCGAGGCAAGAGGG
TGGAATTCCATGAAACAGAGCAAATGAACGAGGTCAGTATCACTGTGGACTATTCC

CAAATTTAGTGTTGCAGTGGATACATTTAGTGGTAACTGCTAATTCTGGTTATATGTG
AGTTCATAAGCCAATGGTGTAGTTCGAATGGGTTAGCCTAATGAGGTAACATAATTA
TTCTAAAGAAATGTATAGCCTAAGTAGCCTAATTAGGTTAGTACTAATGATCTTCTTT
ACCGATCTTCTTTTTTTTTTTCTCGAACACGCAGGAGAACTGCGCGTCATTATATTGA
AAGAGAGAAAGGTCCGAAATGGACCAAAGTACAAAGCCCTAACAGGCAAAAAAAC
AAAGAAAATAAAACAGAGTAGCCTCATACACTAGGAAAAGCCCTAGCCCCCCTAGC
AAACGCAGCCAGAAGCCCAAGACTCTCGAGCTGCTTCGCACCAGCCAAACACCAGC
ACTCCACCTCATCCAAGAAAAGCCTCTGAATGGTGCTTAAAGACGGGTTGACACCAT
CAAATACCGCCCTATTGCGATGCAGCCATAAACACCAGGCCCCCAGGATAATGAGA
CTATGGAAACCTTTACTGATATGTTTGGGGACCCTTCTATGCACCTTGCCCCACCAAT
CCACAAAAGAGCCCGCCTCCCTAGTAGGCGTAAGATGACCCATCCTCAAAGCTGAA
AATATAATATGCCAAAATTGTCTAGCGAAACTGCAAGAGCACAAGAGGTGCAAAAT
AGTTTCCTGCTCTTGATCACATAGTGGACAGACCACCGGGTGAGGTAATCCTTTCCT
CTGAAGTCTATCCGCAGTCCAACATTTATTCCTTATCGCCAGCCAAACAAAAAATTT
GCATTTGGGGGGTGCCCACGACTTCCACACCCATTTCCATGGCTGGAAAATAGTAGC
TCCCATAAACAACGCCGAATAACAAGATTTGGAAGAGAAAACACCAGAGCTCGAGT
GCACCCACACATGTCTGTCCTCCTCCAGATTAAGCACCACTCCACTTAAAGCATCCC
ACAGCTGCAAATACTGATGTAGACCAATCAAAGATAAAGGATTCGTGATATCGCGA
ACCCATTGCCCATTATGCAACGCCTGAGCAACAGTTCTCGAACTAAATGGCTTGCTG
CTCACCATATTAACAACTTCAGGTGCAAAATCCTGTATCCTCTGCCCATTCAGCCAT
CTGTCTGACCAGAAAAAAGTATTGGTCCCATTCCCAACAACAGAGATCACCGAAGC
ATTAAAAAATTGTCTAACTTGCTGCTGCACAGGGATACACAAACCAAGCCACGGCC
GAGTTGGATCAGTCTTCTCTAGCCAAAGCCACTTAGCCTGCAGTGCCCAACTCATCA
ATTGCAAATTAGGAACACCAAGACCCCCTAAACTTAGTGGCCTTGTCACACTCTCCC
AGGGAACAGCACAACTTCCACCATTCACCTCCTTTCGCCCTTTCCAAAGAAACCCTC
TGCGAATCTTGTCAATTGATTTAATAACCCATTTGGGAATTTTAATTGCAATAAGGA
GATAAATTGGGATGGCTGATAAGACAAACCGCACCAATGTTGTCCTGCCGGCCAGG
TTCAATAAACGAGCCTTCCAGTTAGGGAGTCTGTCTGCAATTTTATCTATCCACATCT
TAAGATCATTTCGCTTCAGCTTCCTGTCTGAGATTGGCAGCCCCAAGTAAGAGCAAG
GAAATGAGGCCAACAAGTGGCGTGTTTACCGATCTTCTATTCTTGATATTTTTTGTTC
AATCCAACAAGTGGCGTGTTTACATGCAGAGAGCGACTGCTAAAGGACATGAAGTT
GAAGCTGGATGAAGGAAAAAACTACCTGCAACCGATGGCAATAATGCATTGCAGGT
ACTGGGAATTCACTTGAAATTTTAATGTCTTTTACATTCTTCTATTTTTTGCTGGTGCC
ACAGTGGTAACAGTGTTGGGCTTTCGAAATATTGCTGGGAGATTATTTATTTGTCAA
TGTACCACATGATAGGGTTGAGGCTCGTCAGGAATGTCAAGTTTTTTTTTCTCGAAA
GCGCAGGGGAACTGCGCCTTTAATTATATTAAGAAGAGGAGAAATAGTACATAGTA
CAATAGCTCTCCTTACGGAGGCCAGAAACAAACATAAAGGAACGAATCCATTAGGA
AATCAGACTACCGACAAAGGGGCAGCCAAAAAGGAGAGACCCTTGGCCCCTGCTAG
CACCCAAAGGTTTCTCTCCTCATCAGCCTGCTGCAGGAGTCGAGAGAGGCTGGGGGT
CTCTCCATCGAACACACATCTATTGCGGAGGTTCCATATCTTCCAGGCACCCAACAT
GATGAGAGAATTCAGTCCTTTCTTAACCATTCCCTGCACTGCCTCAGCAGCCCCTTTC
CACCACTCTAAGAAGGATGTAAGGCCAGGCTGTGGGGCCAGGGCATGTAACCCAAA
TTTCCTTAGGAGCATATACCAGAAAATTCTTGTGAAGACGCAGGAGATAAGAAGGT
GGTTCATGGTTTCATCTTCCTGATCACAAAGAAGACACCTTTGCGGGTGATTCAGGC
CCTTTTTGGCCAAACGATCAGCAGTCCAGCATCTGTTGTGTGCCACCAACCACATGA
AAAAACGGTACTTTGAGGGGGCCCAAGATTTCCAGATTCTTTTGTAGTGATCAAAGG

AGGCAGATCCTAAAAACAGACCATTATAAGCTGCTTTCGCTGTGTATTTACCATCTG
CAGCAAGGCTGAAAACATGCTTATCCCTTATGTTTGGCTGTAACTCAATCCCTGAAA
GGGATTCCCAAAGGTTTAGATAATCAACCAAGACTCTTACTGTAAGAGCCCCCTTAA
TATCTGAAATCCATTTCCTGTTTACAAGTGCTTCTTTCACCGTCCTTTTGTTAGCAATT
CTTCTTGGAATTGACTTGTGCAGGCTAGGGGCAATATCAGCGATTCGAGTGCCATGC
AGCCATCTATCATTCCAGAACAAGGTGTTTGAACCATCCCCAACCATAGAAACAAG
AGCCATAGAGAAAAAAGCCCTAGCTTTGTTTGGGACCTGTATTGGAAGACTTGCCCA
AGGCCGCCCCATATCTGTTTTCTGCAACCACAGCCATCTCATGCGGAGAGCCCAACA
AAGCTCAGGGAGACTAGAGATTCCCAAGCCTCCCAGCTGAAGGGGACGACACACCT
TTCCCCAGGCAACCAAACAGTGTCCACCCTTTGCTTCCTTTGTGCCTCTCCAAAGAA
AACCCTTTCTTATCTTATCAATGGCATCTAAGGCCCATTGGGGGATGTCAATTGCCAT
TGCTAAATAAACTGTCATGCCTGTGAGAACATGTTGCACCTGAATTCTCCTTCCTGCT
CTAGACATTAAGTCAGCTTTCCAATTAGGGAGTTGATCTGCAATTCTTTCAACAAAG
GGCTGGAACTGCTGCCGAGAAAGTTTATGCAAGGAAAGGGGTAGCCCCAAATATTT
GCAGGGGAAGGTTGCCATCTCACAAGGCAGCAAACTCTGAACATCCATCAAATTCT
CCTCCGAGCATTGAATAGGATATGCATTTGATTTCTGCACATTATTTATGAGGCCAG
AAGCATCTCCAAAAAGTTGCAGAATGTCTAGAGTGATAGAAGTTTCTTCAGCAGTGG
GATGTAAAAACACAGCCACATCATCTGCATAAATAGAAATCCTGTGGAGCTTCTTTC
TTCTAGATAACTGCTGTAGCAAACCTGCTTCTTCAGCTTTAGCGAATAAGAGGCCAA
GAACATCCATAACAAGGATGAATAGCATGGGGGAGAGGGGGTCTCCCTGCCTGAGC
CCCCTCCTATGGGCTAAACGTTCCCCCGGGGAACCATTGAGCAGCACTTGTGTAGTG
GAGGAGGCCAATAACCTACAAACAATATCCCTCCAGATCTGACCAAAACCAATCTG
CTGCATGACTTCAATTAGGAATGGCCAGGCCACTGAATCAAAGGCCTTAGTGATATC
AAGTTTGAGGAGTAGGCGAGGCAATTTCTGCTGGTGAAACAGCCTTGCAGTTTGTTG
TACCAACATGAAATTATCTTGGATGAACCTTCCCTTTATAAATGCGCTTTGAATTGG
AGAGACCAACTCATTGAGCCTTTTGGCCAGCCGGTTAGCAAGAATCTTGGTGATCAG
TTTTGCAAAACTGTGCACAAGACTAATAGGCCTAAAGTCTTTGACCTGCTCTGCCCC
CTCTTTTTTCGGTATCAAGGTAATAAAAGCCGTGTTGAGCTTGTGGAAATTGTTGAA
ATTCCTGCTCCAAACTGCCCTAACTGCAGCCATAACCTCTGCTTTAATACTTCCCCAG
CAATCTCTATAGAATCGCCCTGTGAATCCATCCGGCCCTGGGGCCTTATCCAGAGGC
AAGCCTTTAATGGTTTCCAAAACCTCTTGTAATAGTTATCTTGGTGTCTTGGAGGTTG
TGGAAACAAAGGAATGCTTGTGTGTTTGATGGGGTATCCCCGAGTATCCGCAGCATG
TTGCAAGAGATTCATGAGGAGGTTAAGCTTTGGGGAATGGCAGGAGCAAAAGATAT
TAGACAATTGTGGCCTTAATTTAAGTGGGGTGGTGGGGGTTGCCCCCCCCCCCTTGT
TTCTAGTTTCTAGTTTTTCTTTCTTTTATTTTGGTTTCTTTCTTCTTTTTCTTTTTTTCT
TTTTGTCCCCCTCTTCTCGTTTGCTGCCTTCTCGGGCACTTTGTTTTCTTAGTCCACTT
GGGACCTTTTTCCTCTTAATATAATGACGCGCAGTTCTCCTGCGTGTTCGAGAAAAA
AAAACCTCTTGTTCTGAAAAAGGGGAGTCTAGATTCTCCAGATTATGTCTAGACAGA
CCAATGGCATCCAAATCAATAGTCCTTTCCCTGTTTCCACACTGCCCAATGAGGTTTG
TATAAAACTGATCCACTGCTGCTGCCTTTTCACAGTGGTTGGTTAGGATACGATCTCC
ATCAACCAGCATGGCCACAAAATTCTTCCTCTTACGATGCCTTGCATGAATATGGAA
AAATCTTGTGTTCGCATCCCCCTCTTTGAGCCATGAAATTCTTGACCTTAATCTTGCC
ATGGTTCGCTTGAAAGAAGACAGCATCAAGGATTGTTTTTTGAGCCTGTTTATAAGC
CAGATCTCTGCTGGGATAAGTGGTCTTTCATCATGTGCCATCTCCAGTCTATGCAGC
AACTCCTTTGCCAAAGCCAACTGCAACCTTACATGCCCAACCTGCTTGTCACTCCAC
CCTTGCAGTCTTCTTGCTGTCATCTTTAGCTTCTTCTCTAGAGTCTGGAAAGGGCATC

TGACAGCTTCGACTGAGTTCCATGCAGAATGCACTTCCTCCTGAAAACCTTCAAGCT
TGGGCCAAAAGGATTCGAAATGGAACCGCCGTTTTCCAGGCCTATTGTCACTGAGAC
CCAGGAGCAGAGGGCAGTGGTCCGAATCAAGTGACGCAGCACTATGAAGAACTCTG
GGGAAGGTTAGCTCCCAATCCACAGTACAAAACACTCTATCAAGCTTAACAAGGAC
CGGGTTATCTTGCTGGTTAGACCAAGTGAATTTGCGTCCTACCAGCGGAATCTCCTT
GAGTCCCAGATCATTGATCAAATCCCTGAACCGACCCATCATCGCTCTATTAAAGTT
GGAGTTATTCTTATCTGATGTTTTATAGATGAGATTAAAATCCCCAGCCACCAACCA
AGGGCCAGCACAGGCTGTTCTGACATCTCTGAGTTCCTGAAGAAAAAGGATTTTTCC
ATCATCTTCCTGAGGGCCGTACACACAGGTAAGCCACCAAGCCTGCCCTTTGCTCCA
TGCACAACTGGACTGAAACACTGTGATTGTCGATTCTTTGCTGCCCTGTATGACCCA
TATGTCTTTTCTGTTGAAGCCTGTGGAAGTCTGTAGAATATGGGTTGAGGTCGGTTC
AGACTTCCTGAACACACCTCCCCCTCTCTCTTTTATTGGCATTTTATGCATTTGGGTC
CTTACAAAATCTTCTAATAACACACACACACACACACCAACACTCCCCCTCAAGATG
AGTGATGAATGTCTATCATCCCCATCTTGCTACATGCTAATAGACACTCTTTAACACC
TAATCCTTTCGTTAAACAATCAGCTACTTGTTCATTAGAGGGTACAGAAATTAGCTT
AAGTGTCCCATCATCCAGACGTTCTTTAATAAAGAAGCGATCGATCTCCACATGCTT
AGTTCGATCATGCTGAACAGGATTGTTAGCAATATTAATTGCTGACTTATTATCACA
CCGAAGCTTCAGAGTACCTTTCATAAGCTTTAATTCTGATAATAAATTCTTTTCCCAC
AACAATTCACTAAGGCATACCGACATAGCTCTGAATTCTGCTTCTGCAGTTGAGCGG
GATACTACTGATTGTTTTTTGCTTCTCCATGATACGAGATTTCCTCCAACAAAAACAC
AGTAACCCGAAGTAGATCTCCTGTCATCCAGACAACTTGCCCAGTCTGCATCACAGT
AACCTTCCACATCTAAATGACAATTTCTTTTGAACCACAGTCCCTTCCCTGGACAGCT
CTTCAAATATCTCAAAATTCTATAGACTGCATCTAAATGTCCACTCCTTGGATCATGC
ATATACCTACTCACAACACTTACTGCATAAGTTATATCTGGCCTTGTGTGGCATAAA
TAAATTAACCGTCCCACAAGCCTTTGATATCTTTCTTTGTCAATAGGATCTCCATTTT
GTGCACATATTTTATGATTTTGCTCAATAGGACTAGAAGCTGGACGACATCCAAGCA
TACCTGTCTCATGTAGAAGATCTAATACATACTTGCGCTGTGAGAGGATAATTCCCT
TATGAGATCTTGCTATCTCAATGCCCAGAAAATAACGTAGTTGTCCTAGGTCTTTAA
CTTCAAATTCCTTGCTTAAATTTTTCTTTAGTTGGGAGATATTCAAAGGATCATTCCC
TGTGATGATCATGTCATCCACATATACTGCCAGAATAGTGACACTGCCTCCAGAATG
ATAATAAAACACAGTGTGATCACTATTGCACTGTTTATATCCCATATTACACATAGC
CCGTCGTAGTCGATCAAACCAAGCTCTAGGTGATTGTTTTAAACCATATAACGATTT
TTTCAGCTTCAATACTTTTCCTTTGGTCTGATCAGTAGCAAATCCCGGTGGGACCTCC
AAGTAGACTTCTTCTTGAAGATCCCCATGTAAAAAAGCATTCTTAACATCCAACTGA
TATAGCGGCCAATCATAATTAGCTGCACAAGAGATTAATGTTCTTACAGTACTCATT
TTTGCAACCGGGGAAAATGTCTCATCATAGTCAATTCCATAGGTCTGACTATATCCT
TTTGCAACTAATCTAGCTTTGTATCTTTCTATTTTTCCATCAGGATTTTGTTTCACTGT
ATATATCCATTTGCAACCAACTATTTTCTTACCCAATGGATAAGGCACAAGATCCCA
AGTTTTGTTTTTTTCTAGAGCAGTTAGTTCTTCAAGCATGGCCTCTTTCCATTTTGGGT
CCAATTTTGCTTCCCTCCAATCTTTTGGAATAGTCACAGATTGTAAGGATGCCAGAA
AAACATTATATTCAGAAGATAGTGAGGTATAAGACACATAATTGCTTATGTCGTGAT
CAAATCCATACCTCTGAGGAGGGACTCCTGCACCAGTTCTTGGTTCCTTTCGTAACG
CAATGGGCATATCATCTGATGGCAGGTCCTCAACCTCAATCTCTGGAGCTCGAGCTT
CTTCAATTTCTTCCTGGCCACTTTCTTGTATAGGAGTTGATTGAGCCATGTCTTCTGT
GTTTGTTCGAGTCATGTACTTCCGTCTTGTATACACTTTTGAGATTACTCTGTCATGT
ATACTTTCTCCTCTAGAATTCTGGGCTTGAGGTAAAATTGATTCATCACTTACCTCTC

TTCTTTCCTGAGGAACTGGACTGCTTATTACTCCTTCATGCTTTGTTTGTCCCTCGTCT
ACATCATCTTGTTCATTCTCATTCTCCCCCTCAGATGTAACTCCATCTCCACTAGGAG
TATCAAGAGTAGCAAATAGATCACTCAGGTCTGATCTTTCTCCATAAAAAGGAATAG
CTTCCCGAAAAGTGACATCTAAACTCACAAACATCCGTCGTTCAGATGGATTCCAAC
ACCTATATCCTTTCTGCCCACAGGAGTACCCAACAAAGATACATTTAACAGCTCGAG
GGTCTAACTTTCCCACTGAAGGTCTGTGATCTCTAACAAAACATGTACATCCAAAAA
CCTTTGGTGGGACAATGAATTCATTCTTTCCATAGAGCAATTCACATGGAGTTTTCAT
GCCAAGTATCCTTGATGGCATACGATTAATTAAATGAGTAGCAGTCATAACTGCCTC
ACTCCATAGAAACTTAGGCACATTCATTGTATACATAAGTGAGCGAGCCACCTCTAA
TAGATGTCGATTCTTCCTCTCAGCTACTCCATTCTGTGGGGGTGTATCGGGGCATGAC
GTCTGATGTAATATCCCTTCCTTAGAGAGAAAAGATCTAAATTCCGTGTTTACATAC
TCTGTTCCATTGTCACTTCTGATAATTTTCACATGAGTATTATACTGGTTCTTTACTAA
TGCACAGAAATCCTTAAAGCAGTCTAACACTTCATTTTTATGCTTCATAAGGTATAA
CCAGGTCATACGAGAATAACAATCAATAAAGGAAACAAAGTATTTCACCCCGCTCA
CTGAGACAATAGGGGATGTCCATACATCTGAATGAACAAGCATGAAGGGACTTTCA
CTTCGTAGTCCTCGAGTCACATAGGAGGTTCTAGTATGTTTCCCAAATTCACACGCA
TCACAGACTAATCTTTGTTTGTTCACCTTACTCATTTCATAAGGAAACATTCTACTCA
TAACATCAAAAGATATATGTCCCAGGCGGCAGTGTAAGAGCATTACCTTTGCTTCTT
CTTCACTGGCCACTGCAGATAGTGCACGACATACTACTTCATCAGTTCTTCTCTTATC
AAGATACCAGAGTCCATTATGACGAACCCCAGTTCCAAGTCTTCTTCCAGTCTTCCTT
TCCTGAATTATACAATTTTCACGATCAAGAAATATTTGACAGTCCAATTGATCAACC
AAAGAACTTATTGAAACTAAATTGACTCCAAAGGAAGGAACATATAATACTGATGA
TAAGGTTATAGATGGTGTGCATTTTACTGTGCCAACTCCTCTGATGGGTTGGAATGTT
CCATCAGCAGTTTGAACCGTTTTCTTATATGAACTTGGATATAACTTATATGATGCAA
ATTCACTTGACCTGCCAGTGACATGTTTAGATGCCCCAGAATCTAATATCCAGCTTG
AATTGGATTTTCTCATGGAAATTAAAGCATGGTTTTTTATACCTGGGTCCAAATTATC
AGTTGAAGTTGACACTGCACCTTGATCATCAGACGTTGAGTCATTCTTATTCTCCTTC
AACTTTCTCAACTCTTTTAGTTCTGAAGCTGGAATTGTAATTAGATCTTCATCTCCCT
CAGGTATAGCAACATTAGCTCTATAACCTCCGCTCCTTCCTCCACGGCTTCCACTTCC
CCTCATAGCACGTCTATCACTGCTTCTTCCTCTCCCACGATAAGGCCTTCGTTGTTGA
GGGCAATCTCGACTCAGATGACCTTGCTCACCACAATTATAGCATGTTCTAGTTTCTT
GTATCCTTGTCATCACAAAAGCTGGAGGAGGTGGAGATGAAGTATCACCTTTTATCA
ACTTCAGTCTAGTATCTTCTTGTGCCATTGCAGCTATGGCTTCTTCGAGACTTGGAAG
ACTGGATTGATGAAACATACTAGCACGCCTTGCCTCAAAATCTGGATTCAAACCTCT
CAAGAATTGAAGAACCCTTCTTTTTTCTATCCATTTCTTTATCCACACTACACATTCA
GTGTGTGGTAGCTCAATAGGATCAAAATGATCCACATCAGCCCACAGCCGCTTCAAA
GCTGCAACATAATCCATTAAGGAGAGGTCCCCCTGCTTCATAGTACTAATTCTGTCC
TCAGTCTCAGCTAATAGCATCACATTACCAGCTCCTGAATACATTTTCGACAAGGCC
TCCCATATTCCAGCTGCACTTGTGATAGTATCCACGGAGCCTGCTATGGTAGGAGAT
AAAGAGCTCAACAACCATGCAGCCACCAGAGAGTTTATAGCCTCCCATGTTTTCCAG
TCATTGCTCGACTTGTCTTCAGGCTCAATTGATTCTCCGGTGACAAAACCTTCAAGTC
TCTTTGCCTTCAGTAATAACATTGCTCTTCTTGACCACGGCAAGTAGTTTCTAATGCC
TTCCAATTTAATGTCATTTGGCATTAGCTCAATTTTCTGGACAGGTTCCATATACGAT
ACAATCTCCTTGCCTGAAGATGGAGTCTCATCTTTCTTCGCTGTAAGCAGTTCCACCA
GCTTCCCAAGGACTTGTTCAATGCCTTGTTGTTCCCCCATTCTGCTCAATCAGCAGAA
GCTCCAGTCACTTTCCTGCAGCTCCTCCGCTGCCTGTTCTTCTCCTCCTTTGCCTTGTG

CCCAGACAGCAGCCAAACCTACAGAGCCGCCCCCTTTTTTCTTGCCGCACTCACCGC
AGACCTCTCTCCGTTGGCCGCCTGCACGGACTTGGTCGTCTTCAGCGGCGCCGCCGC
CGCCTCCTTCGCCTTCTCGGCCGCGGCCTTGGCGTACTCCGCGGTCGTCTCACCGGTG
CTCAGCGTCACCTCCGCCTTCTCCTTGGCGTACTCCGCCACCTTGCGCGCGGTGTCGG
CCGTCACCTCCGTGGCCTTCTGCGCCACCTGCGCGCCCGTGGAGACCGTCGCGTCCT
TGGCCTTCACGGCCGCCTGCTGCGTGTACCCGGCCGTCGTCTTCCCCGTTGCCACCG
CCGCCTCCTTCGCCTTCTCCGCGCCGGGTTCCTTCGCCTGCTGCTGCGCTCGTCCTCG
CCGCCCAGCCGGACGTCGACGTGCTGACCTTGGCCCAGCGTCACCCCTGGCGATGCA
CGCAGGACGAGCCGGCGCACCCCTCCACGCACTTCTCGACGCACTGGTCCACCCGCT
CCGCCGCACGCTGCTGGATCAGCCGGTCCTTCTCCTCGAAGTACACCGACTCGTCGC
CACCACAATCGCCTCCGAGAACCTCGGCTCTGATACCATGTTGAAGCCTGTGGAAGT
CTGTAGAATATGGGTTGAGGTCGGTTCAGACTTCCTGAACACACCTCCCCCTCTCTCT
TTTATTGGCATTTTATGCATTTGGGTCCTTACAAAATCTTCTAATAACACACACACAC
ACACCAACATTTTCCAAGTGATCAAGATGCCACCACTTGCACCTGTTGATGGCAGGT
AGACGTAATCAGTGAACTCAGCACCAAGCATGGAGAGGATAGTTCTTGTTGAAATG
TTCTGCATTTTAGTTTCCTGGACACAAACCACATCCACTTGGGAAGAGACCACAAGC
TCTCTGACTGAGTCTTGCCGCACTGAGGAGTTCATTCCTCTCACATTCCAGAATAAA
ATTTTGTTTGGATTCATGGATCACAGTCTACATAACACAACCACAAAGAACAGGGGG
AGGAACCTAGGCTTCAGACAGGCCCTGCTGGTGGCGTGGGCCACCCGAAAAGAGCG
GACAAGGCCTCCACCTGGAGCTGAGAAAGTGGGTGCTTGAAAAGTTCTGCATAAGC
CTTGAGGGCTTCCTGACTGATGCCTTCATCTTCAGTAATAATCTGCAGGGCTTTCATG
GCTTTTTTGGTGGCCCTGCCACCCCAGTCTTGCATGCTGAATTCCACTCCAGCTCCAG
CGACACGCCGGCTTCGTCTTGGCGCAGCAGCAGATGTTGTTGGAGAGGGAGGAGTT
GCTTTCTTGCTGTTTTTATTTGGTAACAGCAAACAAGCGTCAGGTCTTCTACAAATTT
TGTTGAGGAACTCTTCTTTTCTTGTCGCAACCAGCGGAGCACCCATGGTTGTGAGGG
TATCAGGTCTTGGATCAGTTGTATGCCCATTCTCCAATGCCAGTAACTCCCTTCCTGT
TCCATCTTGTTGGCCGTCGTGAAATTGACGTGCTTTCGATCCTTTGCGGTAATAGTCA
GGAATGTCAAGTTGCTCCTTTTTTTCTCGAACACGCAGGAGAGCTGCGTATCATTAT
ATTAAAGGAAGAAATAAACAAGGTCTAAAATAGACCGGGGTACAACTGACGCCTTA
CGGCGGCCTGAACGAAACTAGCATAGACTGACTCATCTATAAAAACTCTAGGCTAC
ATGAGATAGGGAGGCAACAAGGGATGTCAACCCCTTGGCCCCGGCAGTCTCCCACA
GACGGCGCTCATCCCCCACATGGATTAGGAAATTAGCAACATTAGGGGAAAGGCCA
TCAAAGATGCATCTGTTCCGGAGATTCCACAAAATCCAAGCGCCCAGCATAATGAG
GGAGTTTAGGCCATCTCTGAAGGACCCCTGAATAGCTTCATTTGTTTGATGCCACCA
CTCCATAAAATTAATATCCGTTGGTTGGGGGGCCAAACCCTGCAGCCCGAATTGCCT
TAGAAGGAGGAACCAACATTCTCTTGAGAAAACACAATTCACAAGCAGATGGTCAA
TTGTCTCCCTTTCCTGCTCACATAAAGGACATTTTTCCGGATGATCCAGCCCCCGCTT
CTCAAGTCTATCTGTTGTCCAACATTTCTTGTGAGCCACCAGCCACAAGAAGAATCG
ACATTTGGGCGGAGCCCAAGTTTTCCAAATCCGCTCAAAAGGTTCAAACTCCACAGA
TCCAATGAAGAGCCCCTCATAAGCGGCTTTAGTCGAATATTTGCCATTAGCAGCCAA
CCGAAAAATATGTTTGTCCTCCTTATGGGGGTCCAGATCAACATGTGCCACCAAATC
CCAAATACGCAGGAAGTCTGCAAGTGCACCAACAGATAAACTCCCAGGAATATCAT
TAATCCAGCCATTATCCTCAATCGCCTCCTGAACTGTCCTTTTGTTAATCAGCTTCAT
AGGAATGCCGCTGGAATCTGGTGCGTTGGTTTCGCCCCAGCATACCTTTGCTGGGTC
GCTGGAACTGAGCGTCCAAATTTCGCCCCCTCCAGAGCCTGCTGGGGCCCTCAGTTT
AGTTGCAGCCCAGAATAAAAATTTAACTGGCCAGGGCAGTACGTCCCTCAAAGTTTA

TACTAGGAGGACAGCTGACAACAGGTGGCAAAAGTATGATTCTCAAGCCCTTCAGC
CAAATCAAGCCTCTTCTCCTCAAGAACAGAATGGTATTGGAAATGCAGTCCAAGCTT
TTGTTGACAAAGTAACCAGGAAGGTGGATGGTCTGCTCCCTCCACCACTCCCCCACA
GAAGACGATCCAAGCAACTTCCATCTGACTTCATCCCACGCAGAAGCAGTCGCCTG
GCAAAGCGAGATAATAAAACAAATGATGCACGAAGACAGATTCAAGCTGACCTGAT
AAGAAGCCTGGATGTGCCTTGCGATGGAGAGTTGTTAAGTCCTAAAGCCCTTGAAGC
TTATGGCAACTTGTTCAAAAACTCACTCTCTGAAGCTCATATCAAAGCACTGTCAGC
CCTCTTTGGATGGACAGTCCCAGATGTTTTTGAACCTGTTAATGTGTTGTGTATCGCC
GGGCAAGGCCCAGAAGACCTTGAGTAAGGTTGCTGTTCTAGCAGCACCTAGTGGTTT
TATTTGGTTTTTCCGATGGATCCTTCAAAGGTTCTGTTTTGGAATGTAAGGGGTCTGA
ACAATAAGGCAAGAAGAGACTCAGTGCGTGTTGTGATTGATAACACTAAGCCTGAT
ATTGTTTGTTTGCAAGAAACAAAGAAAGCATCCATTTCTCGGCATATGATTCTCTCTC
TGTTAGGCTCTGATTTTGAAGAGTATGTTGTTCTCCCGGCGGCTGGTACGCGGGGGG
GAATTATTTTGGCTTGGAGGATTTCGGTCTGCAAGATGCTCCAAACTAGAATAGACA
ATTTTTCAGTCTCTGTTCAATTCCAGCAGCAGGATGGTGTGGCTTGGTGGTTTACGGG
TGTTTATGGTCCCCAACTCGACGAGCATAAAATTCAATTTCTTAATGAGCTGCGCCT
GATTCGATCTCAGTGCCTCGGTTCGTGGATGATTGGTGGGGACTTCAACCTAATATA
TAGAGCTGAGGACAAGAGCAATGACAATCTTGATAGAGCTATGATGGGAAGGTTTA
GGCGGCTGCTTAATGACTTGCTTCTAAAGGAACTGCCCCTTTCAGGTAGGAAGTTCA
CCTGGTCCAATGAACGGGCCTCACCTACTTTGGTTCGCCTGGATCGGATTTTCTGTAC
AGCAGATTGGGAGGACATGTTTCCAAACTGTCTCTTGCAGAGCTCTGCATCCCAGAT
TTCTGATCATTGTCCATTACTATTGGGGCTACATGATCTCGTCCAAGGAAAAAGAAG
ATTTCATTTTGAAAGTTTCTGGATCAGATTAGAGGGGTTCTTGGAAGAGGTGGAAAA
CTCTTGGTCACAACCTCTGGAAGTCTGCTGCCCCTTGCAGCGGTTCCATGACAAGCT
AAAAAGGCTTGCAAGGCACCTTCAAGCCTGGAGTCAGAAAAAAGTGGGCAACATTA
AGCAGCAAATTCTGTTGGCAAATGAGATCCTTCACAGGCTGGAAATCGCTCGTGACG
AAAGGAATTTGACTCCCTCTGAGGAATGGCTACGTCGCTCCCTTAAGCACCACTCTT
TGGCTTTAGCCTCCTTGGAAAGAACTGTTGCCCGAGTAAGAAGTGGACTGCACTGGC
TGAAAGAAGGCGATGCCAACACTGCCTATTTTCATCACCACGCCAGACACAGGAAA
AAGAAAAACTTCATTGGAAAACTGATGGTCGATAACAGAGTAATTTCTGAGCAGCA
GGAAAAGAAAGAGGTAATCTGGGATTTCTATAACAAGTTACTTGGGACAGCAGCCC
CCAGGGAGTCAACTTTAAATCTGGAAGAATTTCATCGTCCTGCCCTGGACCTTTCTG
CTCTTGATCAAAATTTCTCTATGGAGGAGGTCTGGGCAGCAATTAAAACTCTCCCTG
CTGACAAAGCTCCAGGGCCGGATGGTTACACCGGAAAATTTTATAAAATATCTTGGC
CAATCATCAAAGAGGATGTTATGGAGGCGTTAAATAGGGTGTTACAAGGTGATGTA
TCCAAACTTCATCTCCTTAACTCAGCTTATGTGACTCTACTGCCCAAGAAAGCTGAA
GCATTGGAAGTCAAAGATTTCAGGCCCATCAGTTTAATTCACAGCTTTGCAAAAATT
GTTACAAAGATTCTGGCTAACAGATTGGCCTCCAGACTTCCTGAGATAGTGTCAGCA
AATCAAAGCGCTTTTATAAAAGGAAGATGCATCCATGACAATTTCATTTTGGTCCAG
CAGACGGCCAAGGCTGTTCAGAAGCAGCGGGTTCCTAGAATTCTTCTGAAATTGGAT
ATTGGTAAAGCCTTTGATTCGATTTCGTGGCCTTTCCTTCTTGAGGTGTTAACGCACC
TTGGTTTTGGTTCTGTATGGTGTAATATCATTTCCAAGATGCTCCTTTCCTCCTCAACT
AGGGTGCTGGTTAATGGAGAGCCAGGGGAAATCATTCACCACCATAGGGGGCTAAG
GCAGGGGGATCCTCTATCCCCAATGTTATTTATTATTGTTATGGATGTGCTGAACTCT
CTTGTTTTAAGAGCCCAAGAATTGAGTCTTTTACAACCAGTGTTAAGAAGAGGAAAT
GGGCAACGTATCTCTCTATATGCTGACGATGTGGTGATGTTCTTGCAGCCAAGTAGA

GAAGAACTCAGTGTGGTCAAAGAAATTCTGAGAATCTTTGGAGATGCTTCAGGACTT
GTTACAAATGTGAGCAAATGCAGTATGACGCCCATTCATTGTGAGGAACAAGACAT
GGTGGCTGCGCAGGAGTTTTTTCCTTGTACTGAGGTGGCTTTTCCATTCAAATACCTG
GGTTTACCCCTTTCAGTCAAGAAACTCCCCAAAAATACTTTTCTAGAACTCATTGAC
AAGATAGCTGATAAACTAACCGGATGGAAATCTGCCCTCATTCATACAGCAGGCCG
GCTGACCTTGGTTAAATCAGTGCTCACAGCCATTCCTATTCATCACCTTATAGTGCTT
CAATGCCCAAAATGGGTGATCAAGGCTATTGACAAAATTCGGAGGGGTTTCCTGTG
GAAAGGTAGAAAAGATATCAGAGGTGGTCACTGCTCTGTTGGTTGGGCAAGAGTTT
GCAGACCGCTTAGCATGGGTGGCCTGGGTATCCATAACCTTGAAGTTCTTGGATGGG
CTTTAAATCTAAGATGGCTTTGGCTCAAAAAAACTCAACCAGAAAGGCCCTGGGCA
GATTTCTTGTTCAAGGTTCACTCCAATGTCGAAGCTCTTTTCTCCGCCTCGGTCCACA
CTACTGTAGGTGATGGCTCTCGAACCCTGTTCTGGACTGATCGCTGGATTCATGGGC
AATCAATATCTACCCTTGCCCCCACTTTGGTCAGCCATATTACAGTGAGAATCAAGA
AAAGCAGATCTGTGCAAGAAGCCTTGGTAAACAATAGATGGGTTAATGACATATCT
GGCAGCCTGTCAGCCCAGATTATCATCGACTTCTTAGTAATTTGGGACCTAGTGCAA
GGTTTTCAGTTACACCCGGGGATTCCTGATGTGCATAGATGGTCCCCGTCTGCATCT
GGAGATTATTCCACTAAATCGGCATATGACCGTTTTTTGGCTGGTTCAGTCTCTTTTG
AGCCGGCTACTAGAATTTGGAAAAGCTGGGCCCCCCCAAGGTGTAAATTCTTCATTT
GGTTGGCTTCTCTAAATCGTTGCTGGACAGCGGATAGATTAAGCAGACGAGGTCTTG
TGCATCCTATAAGGTGCCCCCTTTGTGACCAGCAGGAGGAAACAGTTCAGCATATCT
TAGTGGCTTGTGTGTTTGCCAGAGACACTTGGTGGCAGACTTTGCATAAGGTAGGGC
TGCATCGCCTGACTCCAGGATTGGCAACAACGACCTTCCAAGACTGGTGGAGAGAA
GCTGAGTGCCAGGTGCCAAAAAACCAGAAAAGAGGATTTAATTCCTTAGTGATTCT
GGTGGCTTGGTGGCTATGGAAGCATAGAAATGAATGTGTGTTTGATGGTGCATCCCC
CAACATCAGCATCATCATGCAGCATATTCATGAGGATGCTGTTCGGTGGGGCTTGGC
AGGGGCAAGAGATCTTAGGAGATTGTGGCCTTAGGTGCTGGTGTGGCTGGGGGTTTT
TACTTCTTCTTTTTTCCTGTCTTGATCTATTTTTTTGGCAGCTTGTGCTGGGTCTATTTT
AGACCTCTATCTCCTTTTTCTTAATATATATGATGCGCAGTTCTCCTGCGCGTTCGAT
AAAAAAAAAGCTTCATAGGAATGGTCTCAATAAGTCGGGGAGCGATATCTCCAATC
CGCTGCCCATCCAACCATCGATCACTCCAGAAAAGGGTAGAAGCTCCATTTCCAATC
TCAGAATACAAGGCTACTTGTAGAAACTCTCGAATTTTTTTCGGGGTATGCAAAGGC
AGGGATGCCCAGGGCCGACCTGGCTCAGTTTTTTGCAGCCAAGCCCAACGCATTTTA
AGGGCCCAACTCAGTTCAGTGATGCTGAAAATACCCAGTCCACCAAGCTCTCGAGG
TCTGCACACAGTGCCCCAGGCCACTGCACAATGGCTACCCATGGCATTTTTACGACC
ACGCCAAAGGAACCCTCTTCGAATTTTATCAATGTCCTTCAAAGTAGTAGAAGGCAA
GTCTACAGCCATAACACTATAAATCAGCACCCATGTCAAAACAGATTGCACCAATA
CGCTTCTTCCTGCTTTAGTCAACAAATCAGCTTTCCATCCAGGTAGTTGATTAGCTAT
TTTATCCACAATGGGCTGCACCTGCTCCCTTCTTAATTTATGTAACGACAGAGGCAG
CCCAAGGTACTTGCAAGGAAAGCTAGATATCTCGCAGGGCCAAAACTGTTGAATAA
TCTCTAAATTGTCTTCATGGCATCTGATGGGGTAGACACTAGATTTCTGATTATTATT
ATATAAACCAGAAGCCTCACCAAAAAGCTGAAGGATGTCTGAAATCATAGTCATGT
CCTCTGCCACGGGATTAATAAAAAGAACCACATCATCAGCATAAAGAGAAACCCGG
TGTTGAAGCTGCCTAGTAGAAAGAGGTTGTAATAAACCTTCCTCATCTGCTTTAGAG
ACAAGTTGACCAAGCACATCCATGACAAGCACAAACAGCATCGGAGAAAGAGGAT
CCCCCTGTCGCAGCCCTCTTCTATGAGAAATTATGCTTCCCGGATGACCATTAACTA
ACACTCGAGTAGAGGAAGTGAGAAGCAGCCCTGAAATGATATCTCTCCAGATCTGC

CCAAACCCTAGATTCTGCATGACCTCAATAAGGAAGGGCCAGGAAACCGAATCTAA
AGCTTTAGTAATATCCAGCTTGAGCATCAGGCAAGGATGCCTTTGTTGATGAAGTAA
TTTAGCCGTCTGCTGAACCAACATGAAATTATCCTGGATAAAGCGACCCTTAATGAA
GGCATTTTGGTTTGGAGAGACCAAATCGTTGAGCCTTGAAGCCAATCGAGAAGCCA
ATATCTTTGTGATTAGTTTTGCAAAGCTGTGAACAAGGCTGATAGGTCTAAAATCTT
TGACTTGATCAGCATTATCTTTCTTTGGGATTAGAGTGATAAAAGCAGAATTCAGCT
TATGAAAATTATCAAATTTTCTGTTCCAAATCGCAATCACCGCAGCCATAAGGTCCT
CTTTAATAATGGACCAGCAAATTTTGTAAAATCTACCAGTAAATCCATCCGGTCCAG
GAGATTTATCAGAGGGCAGAGAATTTATAGCATTGCACACTTCATGCTCAGTAATGG
GGCTGTCCAACTCGATCAAATTTGGAAATGGAGTGCCAGCAGTTTCCAGAGCCTCAA
GGTTAATCGAAATATTTCTGTCTTTACTAGCCCCAATTAAATCAAAGTAAAACTGAT
CCACTGCTGCAGCAATGTCGCTATGAGCAGTAAGGACTTGATTTCCAACAGTAAGCT
TGGATACACTGTTTTTCTTCTTTCTGTACCTTGCAAAGGAATGGAACAACTTAGAATT
TGCATCTCCTTCCTTGAGCCAAGAAATTCTTGATCTAAGGCGGGCAATGGTACGTTT
GGTTGAGGCCAAAGCCAGAGCAAGCTTCTTCAATTTACTTAGCAGCCACCTTTCCAA
CACAGACAGAATACGACCATCCTGCGCAATCTCAAGCTGATGGATCAAATCTCTAG
CAAGATGCAGTTGAGTCTCAGTGTGCCCCACCTTCTTGCTGCTCCATCCTTGAAGGG
CTTTGGCTGTCCCTTTGAGTTTTAAATCAAAAGTGGAGAAGGGGCAGTGGCCAACAG
GAACCGAATTCCAGGCGCTGTGCACCACTTCGTGGAACCCTTCAATTTTCGTCCAGA
ATGATTCAAAGTGGAAGCGCCTTTTAACACCTTTGCAATCTTTGAAACCCAATAATA
AAGGACAGTGATCAGAATCTTCAGAAGCCGCACTTTGTAACACTGCACCAGGAAAC
AGCTCCTCCCAATCAATAGAGCACAGGACTCTATCAAGTTTGACAAGAGTAGGCTG
GTTCTGCTGATTAGACCAAGTAAACTTGCGCCCATGGAGTGGAATTTCTTTAAGAGC
GAGATATCTACCAATTTTTTGAATCGACCCATCATAGCCCGATTAATATGTGAGTTG
TTTTATCCTCGGCATGCAAAATGAGATTGAAATCCCCAGTAATCATCCAAGGGCCCT
GACAAGCTGCTCTAATATGTCTCAGCTCATGCAGAAATTCAATCTTGGCTTCATTAG
ATTGAGGTCCATACACACAAGTAAGCCACCACTGGTGTCCCTCCACTGAACAAAACT
GCACAGAAATACTATGGATGTCTGTCCAGTTATCCCCAAGTGTCGTCTCCAAGCAAT
AAGGATTCCACCCCTCGAACCATTAGAAGGCAGGGCCAATGAACCGGTAAAATCAG
ACCCCAACATAGAGAGGAGAACATGCGAAGAAGGTTGTGCCATCTTAGTTTCTTGA
AGACAGATTATATCCGCTCCATAACAATTGAGGAGCTCACGGACAGAATTCTGTCGA
GCTAGAGAATTAAGACCTCTAACATTCCAAACAATTATTTTTTCCGGATCCATGTAG
AAACAATGAGAGGTAAATCGACAAACAATATCAGCAACATCGCCAGCTAAAATAAG
CTTGCTGCGGTAATATCCCCAGTCCCAGTTATATATTCTAAATTATCTGGGACAGAC
CACCCAAAAAGGGCAGCAAGAGCCTGCACATGACTACGAGAAAGAGTATTCTTGAA
GAGCTTGGCGTAGTCGTCTAAAGCTTGACGGCTGAATCCTTCATTGTCCTCAATAAT
ATTTAGAGCCACCATAGCTTGTTTCTTAGATCTTTTTCAAGTTGCTCCTGGACGCGGG
TATGGTGCCCACAAGGCCGACAGAGCTGAGATCAAGCGCTGAGACCATGCCGCCCG
TGCCTGCCTGCGGGCTGCGGCTACATGTCACACATGGCTAGAGGCATAGGCTACCAC
TACTGTTCTCAGAGGACAACCTGCTCGATAGCTGGCCTAGGACGGCCTTTAGTTCCA
GCCAGGCTTCTCGGCCCGTCTCACTGCTGAACGCTATTATTGGCGCCTGATTGAACG
AAGTTGATGTCGGTAGGATCCCATGAAGCAAAGCAACATAAGCAGCAGAAGTCGGA
CGAGGAATGATATACATGCAGACATGCAGTGCCTTCGTTTGTTGCCACTGAAAAAAA
TCTGAACCAAGTATTTCTGTAAGTATACCAACAAGTTAGTTGCTACTACTTTGTATAT
TGATCTTTGTTGGCTGAATGTTTGAGTTTATTTCAAGGCACAAACACTTTATTTTGAG
ATCTGTTCTTCTGTTAGGTTTTACATACTACCTGATTTCAAATGAATTATCTCACGCC

GATAGAGCGCCCGATAGGGCGCTTTCTATTCTAGTTATAATAGAATAGGTAGTGCCA
AAGCATTGCTAAGACAACACAACGAATATTTTGACCATTTTCTGCACCTGAGCAGGC
ACTAATCGGAGTTATATACAATTAAACTCCTGAAATTTTTTTGACACTTACAACTCCA
TTGACAGTTATGCAGGAACAGTATGACCTACATCTGGACATGACACGGCGTGATTTT
CTTGACTGCACATCAAACTGATGTTCTGGATAGCTGATTGTACTGTATAGTTGTTTTT
TGTGTAGGACATTTGTAGCAAGAACTGAACGCTTTTCTATTTGTAGCAAGAACTGAA
CGCTTTTCTATGTACATATTATATATATATATATATAGGGTGATGGTAATTGAAGCCC
TGGGCTTCCATTAGTATGGGAAGCCCCAGTCAAGTATATCTACGCGCGCTCGATCTG
GTTCTGGTTCTGGTTCTGGCTCGAGCGGATTGCAGCGTAAAACGTGACCTAAAACAG
CGTAAATGAGTACGAAATTTAGCGTAAATCTTATATCTGTTTTGTAACAGCAAACGA
GGCCTAAAATTACGGCGTAAAAATCGCATGCAGCCGATCGTGCGCGGGTTCTGGCTC
GGGCGGATTCCAGCGTAAAACGTGAGCTAAAACAGCGTAAATGAGTACGAAATTTA
GCGTAATTCCAAAATCTGTTTTGTAACAGCAAACGGAGCCTAAAATTACGGCGTAAA
AATCGTGTGCACCCGATCGTGCGCGGGTTCTGGATCGGGTGGATTTTAGCGTAAAAC
GTGAACCAAAACAGCGTAAATGAGTACAAATTTTATCGTAAATCATCGATTTGTTTC
GTAACGACAAATGGGTCTCGAATTACGGCATGAAAATCGCCACAGGCGCTGTCTCG
GGCGGATTTCGGCATAAAACGTGAACCAAAACAGCGTAAATGAGTATGAATTTTAG
CGTAAATCATCGATTGATTTCGTAACGGCAAATGGGGACAAATTACGGCGTAAAAA
TCGTGCGCGGGTTCTGGCTCGAGCGGATTTAAGTGTAAATCGTGAACCAAAACAGC
GTAAATGAGTATAAATTTTAGCGTAAATGAATAGAAATATCAGAAACAGACAAATA
GATGGAAGCGGATCCATGAACTGAAATATAGGTCATAAAATCCTATTGACATTGGC
ATAAATATGTATACAAAATAGCATAAAAATAGGATCGCCGTCCTGCGGAACCAGCG
CCTCGCATCTGGGATGCGCCACCACTTCGTGCCTAGGGGAGGATGTCGTCGTCGCTA
GCGCCTAGGGAAATCCGCCGTCGCGACCGCTGGCACATCGCCGTCGTGCGCACCTC
ATGGGGCCGCCGCCGCGTGCGTGCCCTCGGGATCCGATGCCGCAGCGTGTCCCTAG
ATTCGCCGTCGCGCCACAGGAATCCGCCATCGCGAGCGTGTCCCCAGATCCGCCGTC
GCGCCTCAGGGATCCGCTGTCGCGAGCGTGTCCCCAGATCCGTCGTCGCGCCACAGA
TGCTGGAGCGGCGGTGGCGGTGGCCATCGC

>C4contig

GATGATGGCAATGATCAAGGGGGAGATGCAATTGATAAAAAAAAGGAGGATGAGG
AAGAACCAAGACCGCCACACCCAAGAGTCCACCAAGCAATCCAACGAGATCACCCC
GTCGACACCATCCTCGGCGACATTCAAAAGGGGGTAACTACTCGATCTCGGGTTGCT
CATTTTTGTGAACATTACTCTTTTGTTTCCTCTATTGAGCCACACAGGGTAGAGGAAG
CTCTCCAAGATTCGGATTGGGTGGTGGCAATGCAAGAGGAGCTCAACAATTTCACG
AGGAACGAGGTATGGCGTTTAGTTCCACGTCCTAACCAAAATGTTGTAGGAACCAA
ATGGGTCTTCCGCAACAAGCAAGATGAGCATGGTGTGGTGACAAGGAACAAAGCTC
GACTCGTGGCCAAAGGGTATTCACAAGTCGAAGGTTTGGATTTTGGTGAAACCTATG
CACCCGTAGCTAGGCTTGAGTCAATTCGCATATTATTAGCCTATGCTACTTACCATG
GCTTTAAGCTTTATCAAATGGACGTGAAAAGTGCCTTCCTCAATGGACCGATCAAGG
AAGAGGTCTATGTTGAGCAACCGCCCGGCTTTGAAGACAGTGAGTATCCTACCCATG
TCTATAGGCTCTCTAAGGCGCTTTATGGGCTCAAGCAAGCCCCAAGAGCATGGTATG
AATGCCTTAGAGATTTCCTTATTGCAAATGGCTTCAAAGTCGGCAAGGCCGATCCTA
CACTCTTTACTAAAACTCTTGAAAATGACTTGTTTGTATGCCAAATTTATGTTGATGA
TATTATATTTGGGTCTACTAACGAGTCTACATGTGAAGAGTTTAGTAGGATCATGAC
ACAGAAATTCGAGATGTCGATGATGGGGGAGTTGAAGTATTTTCTAAGATTCCAAGT
CAAGCAACTCCAAGAGGGCACCTTCATTAGCCAAACGAAGTACACTCAAGACATTC

TTGCTAAGTTTGGGATGAAGGATGCCAAACCCATCAAGACACCCATGGGAACCAAT
GGGCATCTCGACCTCGACACGGGAGGTAAGTCCGTGGATCAAAAGGTATACCGGTC
GATGATTGGTTCATTGCTTTATTTATGTGCATCTCGACCGGACATTATGCTTTCCGTT
TGCATGTGTGCAAGATTCCAATCCGACCCTAAGGAATCCCACCTTACGGCCGTAAAA
CGAATCTCGAGATATTTGGCTTACACACCTAAGTTTGGGCTTTGGTACCCTCGGGGA
TCCACGTTTGATTTGATTGGTTATTCGGATGCCGATTGGGCGGGGTGTAAAATCAAT
AGGAAGAGCACATCGGGGACTTGCCAGTTCTTGGGAAGATCCTTGGTGTCTTGGGCT
TCAAAGAAGCAAAATTCGGTCGCTCTTTCCACCGCCGAAGCCGAGTACATTGCCGCA
GGTCATTGTTGCGTGCAATTGCTCTGGATGAGGCAAACCCTGCGGGACTACGGTTAC
AAATTAACCAAAGTCCCTTTGCTATGTGATAATGAGAGTGCAATCAAAATGGCCGAC
AATCCCGTCGAGCATAGCCGCACTAAGCACATAGCCATTCGGTATCATTTTCTTAGG
GATCACCAACAAAAGGGGGATATCGAGATTTCTTACATTAATACTAAAGACCAATT
AGCCGATATCTTTACCAAGCCACTTGATGAACAATCTTTTACCAGACTTAGGCATGA
GCTCAATATTCTTGATTCTAGAAATTTCTTCTGCTAGCTTGCACACATAGCTCATTTG
AATACCTTTGATCATATCTCTTTTATATGCTATGACTAATGTGTTTTCAAGTCTATTTC
AAACCAAGTCATAGGTATATTGAAAGGGAATTGGAGTCTTCGGCGAAGAAAAAAAA
AAGGCTTCCACTCCGTAACTCATACTTCGCCATCACTCCAAGCAACTCTCTATTCTCT
GGGGGAGAAATGAGCATCAAAGAAAAGGACTTCATCCTTGGGGGAGAGAGTAAAA
GCTCAAAAGCAAAAGGACCGGATTTCGTCTTTGGTATAATCTTAACTCATTTACTTA
TGACCAAAGGGGAAGAAATTACTTCGAGGGCTCTAATGATTCCGTTTTTGGCGATTC
ATGCCAAAAAGGGGGAGAAATGAGCCCAAAGCAAAAGGACCGCACCACCACATTC
AAAAACTTAGTGTTTTCCAAAAGTATTTATCATTTGATATCCTATTGTGTTCAAAAGG
GGGAGAAAGTAGTATTTCAAAAATGGTATATCAAAACCCTCTTGAACACTAAGAGG
TGGATCTCTTTTAGGGGGAGTTTTGTTAAGTCAAAGGAAAAGCATTTGAAACAGGGG
GAGAAAATTTCAAATCTTGAAAATGCTTTGCAAACTCTTATTCATTTACCTTTGACCA
TTTGCAAAAGATCTTTGAAATAGATTTACAAAAAGAATTTGCAAAAACAAAACATG
TGGTGCAAACGTGGTCCAAAATGTTGTATAAGAAAGAAACATTCCATGCATATCTTG
TAAGTAGTTATATTGGCTCAATTCCAAGCAACCTTTACACTTACATTATGCAAACTA
GTTCAATTATGCACTTCTCTATTTGCTTTGGTTTGTGTTGGCATCAATCACCAAAAAG
GGGGAGATTGAAAGGGAATTAGGCTTACACCTAGTTCCTAAATAATTTTGGTGATTG
AATTGCCCAACACAAATCTTTGGACTAACTAGTTTGCCCAAGTGTATAGATTATACA
GGTGTAAAAGGTTCACACTCAGCCAATAAAAAGACCAAGTTTTGGATTCAATAAAG
GAGCAAAGGGGCAACCGAGGGCACCCCTGGTTTGGCGCACCGGACTGTCCGGTGTG
CCACCGGACAGTGAACAGTACCTGTCCGGTGCACCAGGGGACTCAGACTCAAACTC
TTCACTCTCGGGATTTCTCGGAAGCCGGCGCGCTATAATTCACCGGACTGTCCGGTG
TGCACCGGACATGTCCGGTGCTCCAAGGAAGCGCGGCCTCCAGAACTCGCCAGCCT
CGGGAATGCGCAGCAGCCGCTCCGCTATAATTCACCGGACATGTCCGGTGTGCACC
GGACTGTCCGGTGTAACCACGGAGCAACGGCTATTTCGGCGCCAACGGCTACCTGC
GTTGCATTTAATGCGCGCGCAGCGCGCACAGAGGTCAGGCGCGCCCATACTGGCGC
ACCGGACAATGAACAGAAGATGTCCGGTGTGCACCGGACACCTAGGCGGGCCCAGA
AGTCAGAAGCTCCAACGGTCAGAATCCAACGGCATTGATGACGTGGCGGGGGCACC
GGACATGTCCGGTGTGCACCGGACTGTCCGGTGCGCCATCGAACAGACAGCCTTCC
AACGGCCACTTTTGGTGGTTGGGGCTATAAATACCCCAACCACCCCACCATTCATTG
CATCCAAGTTTTCCACTTCCCAACTACTACAAGAGCTCTAGCATTCAATTCTAGACA
CACCAAAGAGATCAAATCCTCTCCAAATTCCACACAACGCCATAGTGACTAGAGAG
AGTGATTTGCTTGTGTTCTTTCGAGCTCTTGTGCTTGGATTGCTTTCTTCTTTCTTAAT

TCTTTCATTGCGATCAAACTCACTTGTAATTGAGGCAAGAGACACCAAACTTGTGGT
GGTCCTTGTGGGAACTTTGTGTTCCAAGTGATTGAGAAGAGAAAGCTCACTCGGTCC
GAGGGACCGTTTGAGAGAGGGAAGGGTTGAAAGAGACCCGGCCTTTGTGGCCTCCT
CAACGGGGAGTAGGTTTGCAAGAACCGAACCTCGGTAAAACAAATCCGCGTGTCAC
ACTCTTCATTTGTTTGCGATTTGTTTTGCACCCTCTCTCGCGGACTCGTTTATATTTCT
AACACTAACCCGACTTGTAGTTGTGTTTATATTTGTAAATTTCAGTTTCGCCCTATTC
ACCCCCCTCTAGGCGACTATCAGAAACAAAGCCAAACACCATGGTATTCAAGTGGG
TCTTTGGACCGCTTGAGAGGGGTTGAGTGCAACCCTCGTCTGTTGGGACGCCACAAC
GTGGAATAGGCAAGTGTTGGACTTGGCCGAACCACGGGATAAACCACTGTGCCTAT
CTGTGTTGATCCTCTTGTGGTTATTGTGTTTTGCAAAGAACTCATCTCTAGTCACTTT
GCTTTATTGTGCTAACTCCTAATCAAGTTTGGTGGCATTAAGTTTCAAGTTTTTACAG
GATCACCTATTCACCCCCCTTCCCTCTAGGTGCTCTCAATTATTCCCCACAACTTGTT
GAGCGATGAACGTATGTGAGCTCACTCTTGCTGTCTCACACCCCCACAGGTCAAGAA
CAGGTACCGCAGGATGAGGCGCATGGAGGATGCTGCGATGTGTTCGTGAGAGGTCT
AGGTCGTCGTCTCCCAGTCAACTTTGGGTTGCTGGACCGTTGTCTCCTTATAACGTAA
ATATTTATTTATTTTGTACAGAACTCCTATTATATAGTAAAGATGTGACATTCGATCC
TGTGCCATGATTCATCATATGTGTGAGACTTGGTCCCAGCACACCGGGTGATTATTT
CGCGCCCGGGTTTTGGACCCCTAAAACCCGGGTGTGACACCAGGCCCCATGTGTCAG
ACTGACTGACCCGATGCGCGAGTAAGGCGAACCACCGCGCAAGGAGCGTGCAACTG
CCCCACAGTTATGTGCCCCTCTATTCTCGGGGCAACCAGCGGTCAGGAAGGTGAACC
ACCGCACAAAGGAGCATGCAACCGTCCCACGGTTAAACGCTCCCTCATCTTCACCGA
AACCAGCGGTCGAAAAGTAGAACCGCCCCGCGGTTATGCGCCCCCTCAGCTTTGCTG
CAATCGGCGGTCCAACATGGGGGGCCACGCCCACATGTCATGCAACCGGCGCGCCG
GTCACTGGGTGCGGAAAAGTCGCACCGCCACTCGCGCCAATGCCACGACTCCTCGG
GGCCGTCGTGGAAAACTGAAAAGATAAGTTTTCAGAATCAGTGCGGTGACTCGAGG
CACCCGGCGCATGGCCCAGCGGTGTGCGCTAAGTGCGAGAGTCACGGGACGGTCAA
CCGTGAAAACGGGCATGGCCGCCAGCGTGACCATAGGCGGGCCGACAGCAACTGCA
TCAACAGGCTCGTGGGAGCAGCTGGCCAACCGCCAATAAGACTCTAGCCCCACCAG
CAGGGTCGTATCCTCTCTTGAGGCGGGCCCGGGGGCCACTGTCGGTACCCTAAATCA
GGGGTACCCTCTTCTACAGCATGAAGACGCCGCACCTATGTGACGTCTCTAGGCCAC
ACGGAGGACGGTGTCTGGCCCCGCCGCATGGGCAGGCCAAAGGGCGCCACGTGGCA
AGAAAAGACAAGACATCCCAGTATGTATCGTGGGATCCGGACCTCCACAGAAGGGC
ACCAGACCCCTATACGTATAAGTCCGGAGCCTCCGAGTAAGGTCCGAACCTTGACA
AGGTCCCGAAACGAGGAGAACCATGGCGTGAGTAGGGGTCCAGTGCTGACACACGT
CCAGGCCTTGACTTCCCGCTCAGGCGGAGACCCGCTGCTTATGGCCCATGACATAAG
CCATCGGGCCGAGCCTGACGTAAAGTCATGCAACCCCTGCATTTATTGAGGAGGGG
ACGCGTCGCCTGCCACTGCGCTGACGGGCGACGTGCCCTCTCAGCATTTAATGCGTC
CCGTCCCATCCGCTGGCAGGCGATGTCAAGGTCATCCAGCAGGCGGCACGCCTGCT
GGGTCTGCTGGCAAACAGTGCGCCCGTGCCAAGCAGCACACTGTGTTATCCCTTCTA
CAAGAAGCTTCCCCTGCACGCCGAAGATACACAGATCTCGGACGACAGGACACAAG
GAGATTGCCCCGGCAGCAAATATTTATAGTCCCAAGTGTTATATTCTCTATGTCCCTA
GGCCCACATGTCGGGGCTCAGTGCCTTTGTGCATGCCCTCCTTCAGCTATAAAAGGG
GAAACATGCGACATTACAAGACAGATCCAATCTTAAGCTCAGAAACACAATACATT
ACACAGTGGAGTAGGGTGTTACGCCCCGGCGGCCCGAACCACTCTAAACCCTCGTG
TGCTCTCGTGTGTTAATCCACCAACCTCGTAACAAGCAAGACGCTTAGGCCACTTCT
CATTTTAGGAATTAGGGCGGGTGCAATCCGCCACCCGGCCGGATAGATTTACCCTCC

GACATGCACATATATTTTTTTGACAAGTGCTTGTATTTCTCTTTCAGATTACTTGATT
ATGTTGCCATGTGTGATCTGCTAACTAAACTTTGTATGTCAAACTTAATGTTCTTTTG
AGAACACGGATATGAATATACAACACTTCCACATATTGTACCTAAACATTCAAACCG
CTTATTCGCAACTTCAGGTAGTCGCTTCGATCTTTTTTTTCCTCGAATATAATATAGT
TTTAGACATCGAGATACCAAATCAATGGTTTATTGCGATGGTTGGGTTCGAGGGTTA
CTTTAGATAGGTATTATTGATCATTGGTTTTTGTGTGGAATTAGCACTTTCTGATATA
GTTCTTGAGTGTCGCATGCATTGTCTACTTTGCATATTAATCTAAAACAAATAGGAT
AAGGATGTAAAATGACATGTCTGCCTGTAGAACTATGTATGGTTTGTTCCTACTTTG
CCAAATTCATGGCATTTGTGTTCATGATAGAAGGATTCATGTTTCCACTAAGGAGCC
AAAAGTCTGAAGCTACATAATTGCATGACACCATAGTAGGAGTTTATTCAAGTATAA
AATAAGCGATCTGAAGCTGCATAGAAGCACGACACCACAGTTTAGAGGCTGGTTCT
TAACAACTGGAAACCGCTGTTTCATTATGCCAATACCCTTGGTTTCATTGCTCTTACC
TTAGTTCTAACGTTCCCCAACTACTCACAACTTGCTACATTGTCAGCTCCGAGACTTC
ACATTACAGCTTGCTTGGTTATAATTTATGGTTGGGAAATGAATACCACAAGGTTTT
TATTGATGTCTTGGCTTTTAATAATGGCTAGCAAATGACACAATTGCTAACTTGCATT
TTTGTAAGTGTCTATTGTGGATGATGGTGGGGATTCATATACCAAGGAAATCCCTGT
ATCCAACTTCATTTTTCATACTGTGATTTTACCTTTTCCATTCTCTCATTTACCATTTT
AAGCTTCAAATAAAGATGTATTATATATATTATCATATTTTAATTTTTAGTGGTGCTT
ATTGCAGCGGAGTTCATCAACATTGTTCTCTGCACATCAATTGCTTGTTGTAGACACT
TCACGCACGAAAATTAATCCATCATTGTTTTTGTAGCAACAAATATTATCACTTTGTA
TGATGTTTCCAATATCTGGCACATTCATTTGTTGTTATAGGTAAGTTACATCTGTTTT
CTTTTAATGGTTGGACGATGTACGTATACAACTTGAAGGTTCCTCCATGCTTCCACA
GAGACCCTGCATAAAAGAAAGGGGGTTGGTTCAGAAACTTGAACATAGTACAAAGT
TGGTTTGTAAGTTTGTAGCATTTCCAAACAGTTCGGGCTGAGGTATACAACTACTTA
TTTCATCACATGATCACAAAGCAGGAATATGAACAAAGTGACATCTTCGAAAAGAT
ATGAGATGCTAGGAAGTAATTCAAGTCTCGATAGGAGCTTTTGTTCCAACATGACTT
GTGAATCTAAATAAATGATGATAAGGTCGTAAATGCATACTGAAAGTCGCCTAGAG
GGGGGGTGAATAGGGCGAAACTGAAATTTACAAAATTAATCACAACTACAAACCGG
GTTAGCGTTAGAAATATAATCAAGTCCGTGAGAGAGGGTGCAAAACAAATCACAAG
CGAATAAAGAGTGTGACACGCAGATTTGTTTTACCGAGGTTCGGTTCTCGCAAACCT
ACTCCCCGTTGAGGTGGTCACAAAGACCGGGTCTCTTTCAACCCTTTCCCTCTCTCAA
ACGGTCCCTCGGACCGAGTGAGCTTCTTCTTCTCAGTCAAACGGGAACAAACTTCCC
CGCAAGGACCACCACACAATTGGTGTCTCTTGCCTTGGTTACAATTGAGTTGATCGC
AAGAAAGATTGAAAGAAAGCACGATTGCAAAAGCCAAGCGACAAGAGCGACAAAT
AACACACGAATCTCTCTTTCTCTCAAGTCACTAATCACTAATAATCACTTGCCTTAAC
TTTGGAACTTGGAGAGATTGGAGGCTTTGAATGTGTCTTGGAATGGATTGCTAGCTC
TTGTATTGAATGTTGAAGGTTGGAATGCTTGGGTGTTTTGAATGGAGGTGGTTGGGG
TTGTATTTATAGCCACCAACCACTTCCTAGCCGTTTCTCCATTCTGCCGTCCGCGGAC
GGTCCGCCCGCCTGGTCCGGACGGTCCGCGCCTGTAGATCAACGGCTGAAAACGCA
ACGGTCAGCAGTAACGGCTATATTGCATTTAATGCGTCGTCAGATGTCAGATAAAGC
CAGTCACGGACGGTCCGGTCGTGCACCCCGGACGGTCCGCGAGGACGCTATAATTC
ACTTTTCCAAACCCGTCACCTTCGGGTTTTCGGTTCCTCACCTACCGGACGGTCCGCG
CCTGAGGCCGGACGGTCCGCGCGAGGTCTCGGACGGTCCTTGCTTTTCCTTCGAATG
GTCCGTAGTGCAGACTTGGATTTTCGCATTGGTTTTGTCTGAGGGTCATCCTGGTGTT
GCGGACGGTCCGCGCTTGGCCTGTTTTTCTAAAAAGCTTCTCCTGTCCGGAATAATCT
ACGGTATTCCGGACAGTCGATTTAGTATAGTTGTAGATGGACCTTTGGCACCTGTAG

AACATATAATCTAGAGCAAACTAGTTAGTCCAATTATTTGTGTTGGGCAATTCAACC
ACCAAAATCAATTAGGAGATAGGTGTAAGCCTAATTCCCTTTCACATACCATAAGTG
GATTCTCAATCGTAGAATAATGCTACGTAAGATTATCTATGTATTGACGCAAACTGC
AAACTACTCTATCCGTCCCAATGTATCAATATGTATAGTATATAAACATCCACTTTA
ATGATGCTAATAAAAATATAGAATAACAATTATTGTTTTTGCGTATCAATGTAGTTT
ATCATATTAATTCCGTAGCAACGCACGGACATAGGCCTAGTCAATATATAAGGCAG
GACTACTGCCTCCTTGTTTCGAAAAAAGAATAACAGAAACCTAAACCGATATGTATA
AATGTGCTACAATTTAAACAACTAATGGTATGGGTCTAACATCAATTCCAGTAATAG
TCGCAGGATCTGAAAACCACAAATTCAGACTTGGAGTAGATGAATCCAAAACAGTA
ACAACATCCATTGATTCACAATGAGTACCGGCAAGAATATCAGTGCCAATCCTTGAA
CTCTGGGAAGAAATCAGAGATAAAGATAGCAAAGGTTGCTCATAAACAGACAAAAT
CGACGACGCCTGAGGCTTAGGTGTGGAGGTGTCGTGTCGCGAAGGAACAACCACGT
ATGCACCCTCGCCCGATTTGCTTCCACGGTGGGGACTCCGGCGACCGCCGGCGATGG
AGGTGTATCCCACAACTGCGACCCCACGGCTGCGCCACGCCGGACAGAGGGGCAGA
TTAGTCCCAGAGCCCGTGGTTTGAGCGCCCCGCGCAATTAGCGATTTGTAATGTCCT
ACTCATTCGTCAACCGTCTCTTTTTGGCGAGCGCGCCACGTCCTAAACTCCTAATACC
AAAGCAAAGATGACGTGTGCGGGATACCACAGTGTGGTGAAGCCCAGTGATGAGCT
TGTGTCATGCCCCTTGGCCGGCACGGCGTCGCGTGCGGTCGTCGTCGCGAGATCGCC
GATAAACATCGCGAGGTAATTTTGTTAACTGGATTCTGTTACGGCAGACACTTAACC
TCTCCAATAATAACTGACCCAAATACTTCTCATATATAACTCATATATAACGGTACA
AAGCCAATTAACTCCCATACTGGTGGGCTCTTTCGACCTCCGTTGGCCATTCCAATTC
CATCTGAATCGAGAGGCGCCAAACCCCCGCTCCTAAGAACCCGGCGGACGCGCAGC
CCGGAGAGCGCCAGCGCACCGCTGCCGGGACCGGGAGGCAGCCACTCGCGGCCGCG
CCACCGCCGCCGGAGGTCCCCCTTTCCAGTTCCATCCGAATCGAGAGGCGGAGGCG
CCCGGATCCACCGCTCACCCCCGCTCCTCAAACCTCCATGGATCCACCCGCCATGCC
CGCGCGACGGAGGCGCCCGGAGGTTCTGCACGACAAGGAGAACCCGGCGGACGCG
CAGCCCGGAGAGCGCCAGCGCACCGCTGCCGGGACCGGGAGGCAGCCACTCGGGG
CCGCGCCGACGCCGCCTGAGGCCCCCCTTTCCAGTTCCATCCGAATCGnGnGGCCGA
GGCGCCCGGATCCACCGCTCACCCCCGCTCCTCAAACCTCCATGGATCCACCGCTCA
CCCCCGCTCCTCAAACCTCCATAGAnCCACCCGCCATGCCCGCGCGACGGAGGCGCC
CGGAGGTTCTACACGACAAGGAGAACCCAGCGGACGCGCAGCCCGAAGAGCGCCA
GCGCACCGCTGCCGGAACCGGGAGGCAGCCACTTGCGGGCGCGCTGCCGCCGCCGC
CGCCGGAGGTCTGTCGCTCCATCCCTCTCTCCCCGTCCTGGATGAAGTGCTTCAGTTA
GGTAGTGTTTGCTTCCAGATTCTTCTGGTAGGATAGAGCCGCTCCGTTCTAGTAATGT
TAGTAACTAGTCACTGGCGCTCCAGCGCGCGCATATAATCATTGTAATATTAATAGA
AATAACACTAGAGAAATACAAACACTGTGTTGTCACATTGTTACATCTATGTAGGTT
GCTTATCAAAGCATAATAACAAAAGGTGGTTCTCACATATTTACATCTAAGTTGGTC
GCACATGACTGCATAGTAGCAAAAGGTTCGTAGTTGATGGTATGCTATTTATTGAAT
GGAGCGACCATGGAGTATTGGTAGCATCCCAGTCATCCAAATTGCAGCTTGAACCA
AGGTCCTTGCGTCTACAATAGAGAATTAGAGTTAAATATTGTAAGATTAAATGGTAC
AAAATACGATATTACTTTACTGATATGTAATGAAAAATGTGTTTCGGTACCTATTGG
CAAATTGTAAGCATGATAATGTGATTTGTATTTGAGCAGTTTTGTTGGTAGTAGTAA
CGCATTAGAAAAAAAATTATAGTTGTTTAGTGAGGAGGGACAATAAGATATATAGT
TATGTAATAAAAGGAGATCTGGTTTGCAAAAAACAATCAATAACTGGTGTATACCC
GTTATGAGTTTCATATTGCTTTGGTACTGCATGTTGTCGTATTTCAATGGTGAGCACA
GGTAGAAACAAAGGCGAAGTTATAGTTCAAAGCATCAATATGACCTTGGTTTATTCT

TTTTTCTGCTTAATATATAATGGGGCGCAGTTCTCCTGCGCGTTTGAGAAAAAAAAA
TAAGCCAACTATTAAGCAATCATATGTGATACATAAGACATGTTTTAGCGAATAACC
AGGTGGGCGCTATTAGGAAAAAATAGTAAAAATGGAGCACACTGTTATGAAAATGA
ATGTGTAATGAGTTGTATCTTGGTCAATTAGCTTTTTTCCCCTGTGCAGATATGTATA
AGTTGACAAAGCTCATGTTTCGAAACCAAATATTAGGATTTACTGTCAATTTGAAAA
TACCAAAACAGTAAGCAGTCAATGCACGGACGTTTCTGTGCGAAACTACTAAGGTA
CTATAAATACTAAATGAATGTTGTGTTTAGAGGTAATGTAGCAAACCTGTGAAATGT
TTTTCAATAAATGTATCAAGTAATACATATAGATATTAAATGAACGTTGTGTTTAGA
GGTAATGTACCAAACTTGTGAAATGTTTTGAGTAAAAGTACCAAGCAATAGATAGTT
ATGCAAAGAAGAAGTACATACAGAAATAGTTATGCAAATAATAAATACGTAATAGT
ACCACCACTGAAGAGAAATTGTTAAATTGGAAGGCACTGTTAGGGATAGGATTGCG
TAGTGAAATGTACCTTAATCAAGCAGCCTTTTTTTCCCTTGAGTAAATATGTTGAAGT
TGATGAAGCATGTGTTTCCAAAACTAAGACTAGAATTTACTGGCAAATTAAAAAGA
AGGAAACCATAAGAAGTGAATGCATGGATAATTCTATGATGAGATACTATGAACAT
CATGTATACAGAAAATGTAGCTACCACTTGTATTGTTTAGAGTAATTGTAGCAAGCA
TATGGTCTATACGCAGCACGGCACAACAAAATCAGTATCACTACACACCAAAAGAA
TGCTTTCTGACGAAATGTAACTAAAGTTCAAGAGGCTGTGGCTGAGGATGGTATAGC
AAACTTAAGCGAGGACACTGACATGTACGGAGCTCTGAGGGTGAGGATGTTATAGC
TAACTTATTTTTGACCGGACGCCATTAACAATCTGGAAGCACAGGAAGGTCCAACGC
GAAATGAACGGACGACTGATAGCTGCGAGTATGCGGCGAAGCACAGGTAGGGATGT
CAGCAGGCGTCGGCCCGGATGGTCCCTGTCGGCGACGGCGCGGTGAGTATGCGGCG
AAGCAGGGATGATGCTGCGGCGGACGACACCGGTGTAGTGCCAAGGGCAGTGGACG
GTGGAGGAGGGGGGACGTCTAGAGCGGCGACCGACGTCATCTCCGTGGAGAAGATG
GCGGCACGACGGAGACGAAGAAGCCCCTGACTTCTTTCCCATTCTCTTCGATCGATG
ATGTAGAAGTTTAATTAATTCGTATAGCATATGTACGTTCGTATATACAATTAATAT
AGTAAAATAACATTAGTATGCTTTCTCAGCTTGTTATATAATTCAATACCATAGGGA
TCTGCTACACTGTAACCAAACACAGGATGGAGCAGCTCCATTCTACTTCACTATAGA
ACCAAACTGTTCTAGTTACCAAACATAGAACATAACCGTTTTATTTAGAATCTAGAA
TAGAACCGTTCCATTCAACTAGGCTCTAGAACCAAACGCTACCTTAGAACCTTATAC
CGTCGCCTACGTATCTCCTATTACCATCCTCGTGGATGCGTCGATTCAAACTCGGCA
GAATGCATAGATGCCATTTTTTGGAAGGAGAAAAAAAAGACAAATAGTAGCTCACA
ACTCACGTGTTTCTTTTTTCTCGAACGCGCAAGAGACATGCGCATCTTTTCATTAAGT
AGGAAGAGAACAAGATCCAAAATGGATCAGTATCAGTACAAGAAGACCCCCATTAC
GGTGGTCGAAAACAAAGAAACAAAATACATCCATAAATAAAAGGAAACAGAAACT
ACTTTGACCTAAAGACCTGGGATTGGGGCTGTTAAGAGATCTAAGTTTGTAGCCCCT
CCTAACACCCACAGGTCCCTTTCCTCCTCGGTTCTCTTGATGGCTTCATTAAGGCAGG
GAGAAAGGCCGTCAAAAACACAGGCGTTTCGATGTTTCCACAATATCCAGAGGCCG
AGACTGATAAGCGAGTTTAGTCCTTTTCTGGCAAATCCATGCAACTGATCACTACAT
CTTTTCCACCAGAGCATGGTATTCTCTTCATAGACATAAGGAGTGAAATTCTGGAGG
TTAACCTTTGTCAGCAGCTTGAACCAGTAATTTCTTGCAAACACACATCTCACGAGG
AGATGGTCAATGGTTTCCATTTCCTGGTCACAAAGGGGGCACTGATCAGGGTGATCA
AGCCCCCTCCTCTGAAGTCTGTCAGCTGTCCAACATCTCTGTAGCGCCACAAGCCAT
AGGAAAAATTTGCATTTGGGGGGGGACCAGGAGTGCCAAACCCTGTCACAATGCTC
AAAATATGTTGATCCAATAAATAAACTCTCGTAGGCTACTTTGGCAGAGTATTTGCC
ATTATGGGCAAACCTAAAGATGTGTTTGTCCTCTCTATGAGGGTTTATGTTTACTGTG
GTAATAGCATCCCAAAGATCAAGATAATCAGCTAGGACCTCCACAGTGAGAGCTCC

TTGGATATCATTTAGCCAACTTCTATTTGCTAAAGCTGTCTGAACAGTTCTACTGTTT
TGAATTCGCCTTGGCACAGCTAGTAGCAATCTGGGAGCAATATCCATCACATTTCGT
CCATGAATCCATCTATCTTCCCAAAAGAGTGTCGAGTGTCCATTGCCAATTTCCGTGT
GAGTGGCAGCATGAAAGAAGGATTTCACCTTGGAAGGAAAATGCATAGGGAGTCCT
GCCCAAGGTTTATCAGGTTCCACTTTAGTCAGCCACAACCATCTCATTCTCAAAGCC
CAACCAAGTTCTTTAAGGCTGAAAATGCCCAATCCTCCCAACTCAACTGGCCTGCAC
ACTCTACCCCAAGCCACCAGGCAGTGTCCCCCTCTAACTTCTTTTTTTCCACGCCACA
GAAATCCTTTTCTGATTTTATCAATATCCTTGAGAACTCTTGGATGTAAGTCCACTGC
CATTGCAAGGTAGATGATCATCCCTGTCATAACGTATTGGACTTGTACTTTTCTTCCA
GCTCTAGTCATCAAGTCAGCTTTCCACCCCGGGAGCTGGTCAGCAATCTTGTCGATT
ATGTATTGAGTTTGATTTCTGGAAAGCTTGCTCAAAGACAGGGGCAGGCCAAGGTAT
TTACAGGGGAAAGAAGAGATCTCACAGGGGAGACACTCCTGAAGGGTCGCCACTTC
CTCTTCCCCACATCTAATTGGGTACACACTGCTTTTTTGGATATTGGTTTTCAGCCCA
GAAGCTTTTCCAAAGATCTCTAAAATTTTGAGAACCAGTTGAATATCAGATTCTTTT
GGATGCAAAAACATGATCACATCATCAGCATATAAGGACACTCTGTGTTGCAAAGG
TCTAGAAGCCAATGGATGCAGCAGATTTTCTAGATCAGCTTTTGAAATTAACCATCC
TAGCACGTCCATGACAAGTATGAACAACATAGGTGAGAGGGGGTCTCCCTGTCTTA
AACCCTTTTTGAGCTGAATTTTTTCGCCAGCAACACCATTAAGTAGTACTTGTGTGGA
CGAAGTAGATAACAGTCCACACAGAACATTCATCCAAATCTGCCCAAACCCCAGTTT
TTGAAGAACTTCGAGGAGGAAAGGCCAAGATACTGAGTCAAAAGCTTTAGTAATGT
CTAGCTTCAAAAGGATTCTTGCTTGTCTTTGCTGATGGAGAAGTCTGGCAGTCTGTTG
CACAATCATAAAATTGTCTTGGATGAATCGGCCTTTGATGAAAGCACTCTGCCTTTG
GGAGACCATAGATTGCAGTTTATTCCCCAGTCTGTTAGCTAGGATTTTAGTGACTAG
CTTGGCCATGCTATGCACTAAACTAATAGGCCGAAAGTCCTTGACTTGGTCAGCACC
CTCCTTTTTTGGAATAAGGGTGATGAGAGCTGAATTTAATTTGTGAAAGTGACCAAA
CTTCCTATCCCAAATGGCTGCCAAAACTTGCATCACATTCTCTTTAATGATAGGCCA
GCAGGCTTTGTAAAATCTTCCAGTAAACCCATCAGATCTAGGGGCTTTGTCAGACGG
TAGCTCTTTAATGGCTTCAAAGGCTTCTTTTTCTGTGATTGGTAAGTCCAGATCCTCC
AAATTATGAGAGGGCAGCCCCAGGTAGTTAAGATTAATAGTGACAGTTCTATCCAGT
CTTGTTCCCAACAAGCTTCTGTAGAAGTTATCCACCAGCATGGCTTTGTCAGCATGCT
CAGTAAATACATTATCTCCAGAGACCAGTTTTGCCACAAAGTTCTTCTTCTTTCTATA
TCTGGCATGCATATGGAATAGTTTGGTGTTAGCATCACCCTCTCTGAGCCAGTTGATT
CTTGATCTAGTACGTGCAATGGTGCGCTTAAATGAGGCAAGGGCTAAGGAATGCTTT
TTCAGTCCACATCTCAGCCAATCCTCCAAAGGGGACAAAAGCCTGAAATCTTGGGCT
ACTTCAAGCCTGTGAAGAATCTCCCTTGCTAAAGCTAACTGTGAGGTCACATGCCCA
ATCTTTTGTCCACTCCATTTCTGCAAGTATCTAGCAGTGGCTTTTAGTCTCCTATCAA
AGCTTAGAACTGGACAGACTCCCTCATCAATAGAAATCCAAGCATCATTGACCACTT
CTTGGAATCCATCTAACTTTGGCCAAAAGGATTCAAAATGGAACCGCCTCTTGCCAA
AACAGTTGCTCTGCAGACCTAGCACAAGGGGGCAGTGATCAGAGTCCTCCGAGGCA
GCGCTTTGGAGGAGGTTATTGGGAAAAATATCTTCCCAGTCACATGAGCACAGGACT
CTGTCTAATCTAACCAATGTAGGAGACACTTGGTGATTAGACCAGGTATATTTGCGC
CCTATTAGAGGCAGGTCAATAAGGCTTGTATCATCAATAAATTTCCTGAATCTCCCC
ATCATGGCTCTATTGCAAAAAAGCGTTGTTTTTGTCATCCTCTTTATATATAAGATTG
AAGTCCCCTAAAACAACCCAAGGCCCCTGACACTCTGCTCTGATCTGTCTGAGTTCT
AGCAGGAAGGAAATCTTGTCTTCATTGCTTTGGGGACCATAAACATTTGTTAACCAC
CAATGCTGGCCATTGGAGTTTTTGAACTGCACAGATACGCTGAAATTATCCACTCTA

CTCAACCCTGTGACTTGTAAATGGTTCTTCCAGGCTACAAGAATTCCTCCGCTGGCA
CCCAAAGAGGGTAAGTAGACATACTGCTCGAAGTTTACTCCCAGAATGGCTATCACT
TTTATTCGGGTAATTTCCTGCATCTTGGTCTCTTGTAAACAAACAATGTCCACTCCCA
GGGAATCAATAAAACTGCGAACAGAACTCTGTCTCGAAGTAGAATTCATGCCCCTA
ACATTCCATATAAATATTTTTTTTGGATCCATAAATTACATGTTCAGAAGAACAACA
AAAGAGAGAAGTGGACACTCAGCAGGCAGCAGCACTATGCGATGAAGCCTCTGCCC
GCAGCTCGTCCGGCACCTTCCACCCAAAAAAGGCTGCGAGAGCCTCTATGTGTTTGA
GGGGAAGTCTTTCAGAGAATATTTTTTCATAATCTTTTCTGGTTTGCTGATCTATCTC
TTCCAGTTCCCCCACCACCTTGAGAGCCTTCATAATTTTCTTTGTTGCTCTTGAGGAA
AGGTCTTGCTTGCCAATTTCTGCTCCCATTCCAGCAATGCGGCGGCTGCGACGAAGG
GGCACAGCAGTGGACCACTCAGAGATTATGGGCTTTGGAGCTGGATTTGCTGGTGCT
GGCAGAATACCAGAGACTTCTTTCACAATTTGTTTAAGAAATACCTCTTTTTTGGCTG
TTACAGGGTCCGTGCCTTCCATCTGCAGGGCTGGAATGATATTATTTTCCGTTGTTTC
TTCAGCAATTTCAACTAGCGGCTTATTATGCTTCTTTTTTCTATGATAAACGAGCAGG
CTTGGTGATAGTCGCCTAGAGGGGGGGTGAATAGGGCGAAACTGAAATTTACAAAT
ATAAACACAACTACAAGCCGGGTTAGCGTTAGAAATATAAACGAGTCCGAGAGAGA
GGGCGTAAAACAAATCCCAAGCGAATAAGCAAGTAAGACACGGAGATTTGTTTTAC
CGAGGTTCGGTTCTTGCAAACCTACTCCCCGTTGAGGAGGCCACAAAGGCCGGGTCT
CTTTCAACCCTTCCCTCTCTCAAACGGTCCCTCGGACCGAGTGAGCTTCTCTTCTCAA
ATCAAAGCCGGGAACAAAACTTCCCCGCAAGGGCCACCACACAATTGGTGCCTCTT
GCCTTGATTACAATGGAGTTGTGATCTCAAGAACAAGTGAGAAAGAAAAGAAGCAA
TCCAAGCGCAAGAGCTCAAATGAACACGGCAAATCACTCTCACTAGTCACTAGGGT
TTTGTGTGGAATTGGAGAGGATTTGATCTCTTTGTATGTGTCTAGAATTGAATGCTAG
AGCTCTTGTAATAGTTGAGAAGTGGAAAACTTGGATGCAATGAATGGTGGGGTGGT
TGGGGTATTTATAGCCCCAACCACCAAACTTGACCGTTGGCTGGAGGCGTCTGCTCG
ATGGCGCACCGGACAGTCCGGTGCACACCGGACAGTCCGGTGCCCCTGCCACGTCA
TCACTGCCGTTGGATTCTGACCGTTGGAGCTTCTGACTTGTGGGCCCTCCTGGGTGTC
CGGTGCACACCGGACATGTACTGTTTGATGTCCGGTGCACCGGCATGGGCAAGTCTG
ACGTCTGCGCGCGCTGCGCGCGCATTAAATGCACCGCAGGGAGCCGTTGGCGCCGA
AAAGAGCCGTTGCTCCGCTGGTACACCGGACAGTCCGGTGCACACCGGACAGTCCG
GTGAATTTTAGCGGAGCGGCTGCCATGCGAACCCGAGGCTGGCGAGTTCCGGAGAC
CGCGCTTCCTTGGAGCACCGGACATGTCCGGTGCACACCGGACAGTCCGGTGAATTA
TAGCGCGCCGGCTTCCGAGAAATCCCGAGAGTGAAGAGTTTGAGTCTGAGTCCCCTG
GTGCACCGGACAGGTACTGTTCACTGTCCGGTGGCACACCGGACAGTCCGGTGCGC
CAAACCAGGGGTGCCCTCGGTTGCCCCTTTGCTCCTTTATTGAATCCAAAACTTGGTC
TTTTTATTGGCTGAGTGTGAACCTTTTACACCTGTATAATCTATACACTTGGGCAAAC
TAGTTAGTCCAAAGATTTGTGTTGGGCAATTCAATCACCAAAATTATTTAGGAACTA
GGTGTAAGCCTAATTCCCTTTCAATCTCCCCCTTTTTGGTGATTGATGCCAACACAAA
CCAAAGCAAATAGAGAAGTGCATAATTGAACTAGTTTGCATAATGTAAGTGTAAAG
GTTGCTTGGAATTGAGCCAATATAACTACTTACAAGATATGCATGGAATGTTTCTTT
CTTATACAACATTTTGGACCACGTTTGCACCACATGTTTTGTTTTTGCAAATTCTTTTT
GTAAATCTATTTCAAAGATCTTTTGCAAATGGTCAAAGGTAAATGAATAAGAGTTTG
CAAAGCATTTTCAAGATTTGAAATTTTCTCCCCCTGTTTCAAATGCTTTTCCTTTGAC
TTAACAAAACTCCCCCTAAAAGAGATCCACCTCTTAGTGTTCAAGAGGGTTTTGATA
TACCATTTTTGAAATACTACTTTCTCCCCCTTTTGAACACAATAGGATATCAAATGAT
AAATACTTTTGGAAAACACTAAGTTTTTGAATGTGGTGGTGCGGTCCTTTTGCTTTGG

GCTCATTTCTCCCCCTTTTTGGCATGAATCGCCAAAAACGGAATCATTAGAGCCCTC
GAAGTAATTTCTTCCCCTTTGGTCATAAGTAAATGAGTTAAGATTATACCAAAGACG
AAATCCGGTCCTTTTGCTTTTGAGCTTTTACTCTCTCCCCCAAGGATGAAGTCCTTTT
CTTTGAnGCTCATTTCTCCCCCAGAGAATAGAGAGTTGCTTGGAGTGATGGCGAAGT
ATGAGTTACGGAGTGGAAGCCTTTTTTTTTTCTTCGCCGAAGACTCCAATTCCCTTTC
AATATACCTATGACTTGGTTTGAAATAGACTTGAAAACACATTAGTCATAGCATATA
AAAGAGATATGATCAAAGGTATTCAAATGAGCTATGTGTGCAAGCTAGCAGAAGAA
ATTTCTAGAATCAAGAATATTGAGCTCATGCCTAAGTCTGGTAAAAGATTGTTCATC
AAGTGGCTTGGTAAAGATATCGGCTAATTGGTCTTTAGTATTAATGTAAGAAATCTC
GATATCCCCCTTTTGTTGGTGATCCCTAAGAAAATGATACCGAATGGCTATGTGCTT
AGTGCGGCTATGCTCGACGGGATTGTCGGCCATTTTGATTGCACTCTCATTATCACAT
AGCAAAGGGACTTTGGTTAATTTGTAACCGTAGTCCCGCAGGGTTTGCCTCATCCAG
AGCAATTGCACGCAACAATGACCTGCGGCAATGTACTCGGCTTCGGCGGTGGAAAG
AGCGACCGAATTTTGCTTCTTTGAAGCCCAAGACACCAAGGATCTTCCCAAGATnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnCGGGAAGGGAGGGGAAGG
GACGTACCTGACGGTCCATGTTCTAACTGCTTTAACTCACGTGTTTCTCTTCATGAGT
TCCTCTTCATGAGCTCATGTGAACTGGGCGTGTTAATTTCTGGCATATGCTCATTCTC
TAGTTCAACTCTGGACACTAGGTGATAGATGCCTGTTAATGAATAATGCAATCCTTA
TTTGGTGTTGTTTAATTGAAATATCCTGATGTATTGATTATGCTGCTGTCCTTCTAGT
CTTTGCAAAATGAAGCTGATTTATTGAAGGAGGTTGAAAGTCTTAAAATCGAGCTTC
AACATGTAAGGCAGGACCGTGACAGGTTACTGTCTTATATGGGAACATATAAAGAA
ATGACTGAAAAATCTGTCATTGAATTAGAAAACGCCATGAAAAAAGCCTCTACCCTT
CAGGTATGACTATGTGTCTAAATAATTTGATCGTTATCTAGATAAATGAAACTGGCT
AACAAGTTTTCAATGACATAGTGTTAGACTATGTGTGTGTGAGTGTGTATGGGTCGG
CATCACTGTTGGGCTGCCGGCCCATTAGGGTTAGGGTTTCCTGTGTCTCTATATATTG
TAACCCCATCTATTATCAATACAACACTTTACACTTCTACATGGTATCAGACTAGGTT
AGGGTTTCTTCCTCCTCCCACCCACAGCCGCCACCCAGCCCTACGACCAGCCGCCGC
CCCCAGGAGGCTCATCCCTCCTCCCAGCCGCTGCAGCAGAGCCCCCTGCGAGCAGCT
CCTCTAGGGAGCACGGTCTCCCTCCAGCCGCTCCTCCCCATCTCTCGGCGGCGGCCC
TGACAGCGCGGGCGGCGTGGGCATCGGCGACATGGGCACGCGGCGCGGCGGCCAG
GCCCCTGCGTGGGTGCGGCTGCGCGGCCTCGGTTGCTCCGGTCCTAGCACGCGGCGG
CGCAAGCCCGTGTTCGCGCGTTCCCCTGCGCGGGGCGGCCACGGCGAGCGGCCCCG
GCGCGGGCGCGCGTTACCCTGCGCGGGGTGGCCACGGCGCGGGCTGGCTCGACGGC
GGCCATGGCGCACGCGGGCAGCGACCCTGCGTCCCCTCAACGTCGGCCTTCCCCCTG
CTCTTCCCGGCCCTTGGCCAGGGAGCGCCCAGGTCCGGCGACCTCCCAACTCCATGG
ACGCAGGGACCTCCTTCCTCCTCCATCCAGCGCCCCCTCCTTTCTCCCAGCCGCTGGA
Cnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnCCTCATGCCGGCCAGCCCCTC
CTCTTCCCTGGCTGCACGCATGTCGCCTCACTGGCCGTGGGCGCCCCCCCTTCCCCG
GCCTCCCAAGCAGCGCCACCGGATCCGCGCGCCCCTGTGTCTGCGTCGCTTGATTCG
CGAGCCCGCGGGCGCGCGCCGGCGTTCTCTCCCTAGTGTCGATCACCTCGACCAGCA

GCACCTTGTAGCACCACCCTTCGAACTTCTCCCGTCACGGTGCCGTCAACACCCCTT
GCCTGGTCCGGCGTAGGCCCCTGCTCGGCTGGACCCGCCCCGTCGCCCTACCCCTTC
ATCGCCGTCATCGTTGGTTCCGTCGCTGCCGCGGCCTTCCCGGCCGGATCCACGTTG
CTACGGCCTTCTCGGCTAGATCCGCCACCCTTTTGGCCGGATCGCCGTCTTCATGGCT
GGATTCGTCGCTGCCGTGGCCTTCCCGGCCGGATCCACCGATACCCCTTCTACTCTG
CAGCGGGCGCGGACACCCTGACCGGCGCAGTCGCGTGCGCTGTCGTCGGCCTCCCT
GGCGGGGCTCCTTGACGCCTGGACGATCGAAGAAGCAAGAAGGTCGGCCTGCCGGT
GCTCTTTTGTTGTGTCGCCCTGCCGCTCGTTGACGCTCGAACGTCGACCCCTCCCGAA
GATCAGGCTTTTGCTGCGCGTCTTCCGACCACGACCACCTCGACTTCGGCTACCTCG
GCATCTAGGGGCTATCGTCTTCTTGGAGCACACACCGGTCTCTACTCCAGCCGCAAC
ATTCGCACCATCACGACGTTGCGACTGCGGGGGGATTTCAACCCGTCGGCTCCTACC
TTCGGCCTCTACTCCAGTCTCATCGTGTGTGGTGCCCCCGTTGCGACTGCGGGGGGA
TGTTAGACTATGTGTGTGTGAGTGTGTGTGGGTCGGCACCACTGTTGGGCTGCCGGC
CCATTAGGGTTAGGGTTTCCTGTGTCTCTATATATTGTAACCCCATCTATTATCAATA
CAACACTTTACACTTCTACACATAGGAAACATGTTCGTCCCAGAGAAATATTATAAG
AACATTGGAAAGTAAGCTTGCAACTGTGGAGAAGTTAAAGGTGCGTACATACCTAT
TTTTTTCATTAATCATTACATGTAATGAACTAATGGTGTTGTGTTATATCTCAGTAAA
CACAAGTTTATTTGTATGTTTCTTAACGTATGAAGTGTTCACATTGTGTAGGAAAAAT
ATTGCATAGATAATTAATATCTCGAACCTTCTCTTAAATGTACAGATGCCTAATATG
ATGTCCTCAGAAACCATGACTGAGTATGAGGACATGAAGAAATTGCTAGAGAGTGT
CCAGTTACGCCTTAAAGAGGCTGAGCAAACAATTTTGAATGGAGAGATTTTACGCA
AAAAGTTACATAACACAATTCTTGTAAGAATTCTTTTAAACTCAAAAATCTAGTATA
TGTACACATATGCTATCCGGCTATTCCTGAATTTGTTTTTTTTTGCAGGAGTTGAAGG
GAAACATCAGAGTATTTTGTAGAGTAAGGCCTTTGTTGCCAAATGAATCTGGAGCGG
TCTCATATCCAAGGAATATACAAGACATAGGCCGAGGCATTGAGTTGTTACATAATG
GTAGAGTCATCTTTACAATTATTTTCGTAGTTTTAGACTTTGATATGTAGTATTTTAG
AATATTGATTATGATATATTTGCATTTCTTATTTTTTCAGCACAAGGATATTCATTTA
CTTTTGATAAAGTATTTGATCATTCCGCAACGCAGGAACACGTATTCACAGAAATAT
CTCAACTCGTGCAAAGCGCTCTAGATGGCTATAAGGTGCAACATTAGGATGACTATA
ACTTAGATTACCTTTGCAAACTCTAGAAATAATAATTTGTTGGTTTTTTTTCCAATAT
CTAGTTTTCTTAGAATGCAACTAAAAAAATGGTAAATCATTTTAGGTCTGTATATTTG
CCTATGGGCAAACTGGTTCTGGTAAAACATACACAATGATGGGCAACCCTGAGTTA
GAAGAACAAAAAGGAATGATACCACGATCTCTGGAGCAAATTTTTCAAGCTACCCA
GACACTTAACTCACAGGGATGGATATATAAGATACAGGTGCTTTTTTTTGGTAATTT
ACTTATTTAATTGAACTGACCAGTGAACAGGTTTTTTGTAACTGTTATCATGGTGTAT
TTTAAACTTTCAGGCCTCAATGTTAGAAATTTATCTGGAAACCATACGTGACTTGCTT
GCGACGAAACACATTTCTCAAGATGGTGGTTCTTCAAAGCACAATATCAAGCATGAC
GCACAGGGTAGTATTGTGTCGGGCCTTACGGTCATTGACGTAACAAGTATAAGCGA
AGTTTCTTCTCTCCTTAGGCGGGCTGCTCAGAGCAGGTTACACGCTTATTTATTTTCA
AAGTTTCAGCATTAAGTATTGGATCGCAGGAGTATATAATAGCAAAATCTTGTACTG
ATCACGAATGTTTATGTATGTGCCATAATTTTTGGCAGCAGATGTAGTATATAATGT
AGCTATGTAAAGAAGTCACTACATCATATTCCCTTGACTGTCATATCACTGCCACTG
CGTTCTCAAACTTCTTTTGGAGTTCTAGATTTTGAGAATTGAGTGGTGTAGGACAGTC
ATCTATCAATTTTGTTTCCAGATCAGTTGGGAGAACACAAATGAATGAGGAATCATC
CAGAAGTCACTGTGTTTTAACTCTACGGATTTTTGGTGTAAACGAGGTATATTCTAAT
TCACATGTATCGTTGTTTGGCATGTTGTCTTTTCGTCTATAACTGACCGTTTTTTCTCT

TCAGAGAATGGACCAACAAGTGCAAGGAGTGCTCAATCTTATAGATCTAGCAGGCA
GTGAGCGCCTGAACAAAAGTGGTGCCACTGGTGACCAGTTGAAAGAGGCTGTGGTA
TTTTGTTATCTAGTCCCTTAGCATTGTTTGTGTTTTAGCTTGATATATATAAATAAAA
TAACGAAGAGGCATGGTTTGTTCCTACTTAGGCCATTAATAAAAGCTTATCATGTTT
AAGTGATGTCATCTTCTCAATCGCAAAAAAAGATGAACACGTTCCGTTTCGGAACTC
GAAGCTGACCTATCTGCTCCAGGTATTCCCTTATCCTCTTGCATGCCATCTTGTTTTC
GTGAAAGTCAAATTGTAAATGTATCAACTGATCTCCGCTTCCCAACACAGCCATGTC
TCGGAGGGGACTCCAAGACGCTGATGTTTGTGAACCTATCCCCTGAAGTGTCTTCGA
CTGGGGAGTCTCTCTGCTCATTGCGGTTTGCTGCGAGGGTGAACTCGTGCGAGATCG
GGATCCCTCGGCGCCAAACCAGAATGTCGTCTGACCGACCAAGTTGAAATAGGACA
TCAAGTACTCAGTAAGTCATCATTGAACACCAAGTCACCAACGCAGGAGCTATTAG
AATCATGTAATGAATTCAAGTATGTTAATAATTATCTCACATTGACTGCATCCACAT
ATTAACCGATATCTGGACTAATACAACAAATCTGGAATATCATTATATTTCTTTGAG
TGTTATTCTGACTGGGTCAGGTGGCGCCCTTTCTGGTGCCCTATCCAACACAAGGGG
AAAACAGTAATGCAGAGGATCTGAGAAATTATTCAGCACAATTGATGTTTTAAAGG
TTGAATGCCCCACAAAATTGGTTAGCGGCGGGAGACAGCTAGAGCAACAAATAAAA
CATAGTAACAAACACAGAGAGAGCGAGACTGATCATAATTCGCATTAAAGTTGCAA
CTGTTTCTTGTTTCAACACAAGCCACATACATAGAGGTGTAGCATAGGCCCTCTACA
ACAGGGAACCGACCAAAATAAGGTGACAAGGTCCGTACGTAGGGCCAAACACCAC
AACAGGAGGTATGTACATGATGCTGGTAGGGGTGGTAATGGATCACGATCCAAATG
TTTGTCCACAATTTGTTAGGGTCCTAAATTAATTTTAGTTCAAAAATGAATAGAAAT
AGAGCCCGATCCTAACCTGATCTGATCCTTAAATTTTAGCGTAAAATTTAGATTCCA
TTGTCAACCCTAGGTACTGTAAGACATACTCTAGCTAATGCAGTCCACATAGTAATG
TACATAAACTTTACCAAACCACAAGAATGTGAGTATAAAAGTGCCATAGATATTCCC
CGTGATGCGCTTTTCTTCTCCAACCCTGCCTAATTACTTCTAGTTAAAAGATACACGG
TTTTGCTGGCTTCTGAGCGATTCACGCGGATTCGGTGTGTTGCGCGGCGACGCGTGC
CGTGGTGGGGTCGTGTTTGGCGGGCTTCGCGCCATCCGTTGCGCGACGCGCGTCCTC
CGTGGGTCCGGCGCACGCGGATGTGAGACCTCGCTTTCGACGGACGCGAAATTTCC
AGCCACACACCTTCTCTCGCAGTACAGACGTACAGTGCAGGCATGGTCGCGGCTGTT
CTTAGCTGATGGCGAACTGGCGAAAAGGCCACGATGCGGGACCCGTCTCTGCTCCA
TGCCTCCATTACTTTGTGCCGTCCTCCCTCCTTCCACGCTCCTTCCCGCAAGAGCAGC
TTCCCTCCTCTCCTTCCCCCAAACGCTCACCTCCTCTTCTCTCGCGCTGAACTCGACC
TAGCGTTCGACTCCCTCGCGGCCCCATCGCAGGGTGTCGTCGTCCCTGCCACAGTCC
GTCGGGGTCGGCTTGTCCTCGGATCTCGCACGATTTAGGAGCAGCGACATCAGATGT
TACGGACGCGAGGCCATCAACTCCGCCGCCACCAGCGCCGATGCCGGCCCCGGCCC
GGCGGATCCCTCCCCGGTGGCCCCGGAGGCTACAACCTGGCAGATGTCGGAGGTGT
GGAAGGAGATCTTCGCCTCCGTGCCCGTCCAGTTCCGCTTTTCGCTTCTTCAACAGG
GAAAATTTCCTTTCTATTGTCACCCCCTGATCGTCCAGTTTCGCTTAAGTTTCCTTGCT
ATTGTTAGCAGGAGGCTGTGGCCCGACATACGCTAGAGGATTTCCTGCGCCAGGGG
GGTGGTGATGGAGGGGAACATGGCACTGGCCTGTCACATATTTGCCACCTGCAAAT
GGCTGATTCTGATTTAATCACAAGGAAATAATATGGGTTGTCATTATTTGGATTATG
TCTTAATGTCCTGTACATGCTGGATCAGGTTGACCTTGATCCCTATGGGTCACTAGCT
GCATTCCTTGATTCCATGGTTCAATGTGTAGCAGATGGGGGTACGCTAATGTGCTCA
GCTATAGACATGATAGTGCTTGCTGGTGGAAATGCAGAAGTTTGTTTTTCTAAATGA
GAGTTTCAGTTTTAGGATTAAATACGAATTGCTATCTGCTACACATTGTTTTCCTAAA
ACCTTTTGCTTCCAGCGTCTGCCTCCACAATACCAACTTTTGCGAAACACCACTCCTA

CTCTCCTCAATAAGTACCACATCATCGGCAAAAAGCATACACCACGGTATATCTCCT
TGTATGTCCATGTAAAAATTTAGTGTCTTCAAAGTTGTCCTTTACTGATTGTTTGGAC
AAGAAAAATGCCAGTATTTTTTAGGAGAATTCCAAACTATGATCTACGAAAACAAA
ATAAAAAGGTCTCCAGTTTATGTTCATAATTTAACCTTTTTCTAATATATGAAATATG
AAACACAACTAAATGCCATATGCAGTAAGTTCCTCCTGACATCATGTATTATTTGAG
CTATAACTACTTTGTTCTTGACGTAGATAAAAGAGTTAAGTTGCTTGCAATAAACTC
ACAACCTGTGTTATTTCAAGTTACAAAAAAACCATGTTACAGGTGAAATATGTATCC
AATTCAAAAGTATCATAAATAGCAAATCCAGCAAAGATCCTGTGTAAACTGGACTG
TCCTTGCAAGCTCTTGATAGCCACATAAGTGTGAACCAGATAACAAACTATTATTCT
ACATGCAGTTCAATAGGAAGCATGGTCACCTTTCGTAAAGAAGTAAAGACTATAGTT
CATCCCGGTGAGGTCTCAGTTTTAGGTTCCCTTCGAAGTATATCACCTGCTATTGGGT
TAACCTAGGGTGCATGCAGACAACCAGTTGCACTGTGGGATCTAATCATTTGTGTTT
ATGGTGCCGATTTTGGTTCCTCATTCTCGCTTGGTTTATCTGGGAGGTCTGCTCCCTG
AGAGAAAGAACAAAAACAAAAGGAACCTAAAAGAATTTCTGTTATTACATCCACCG
TATGGGTTGTAATATACTATGTCTGTTTTTTTGGTTCATTGGTTTTGAAATTCTACTAT
GATATCTATACATAGTAAGTGAAATGTTGAGCCTTCAGGACATCGGTTGCTACCACC
TGAGGCTTTAAACTAATCTCAATCACTTCTGCTTTTGCATATTGGTTTGGGTGCATGA
CATAATTGGTCAAAGCAATGCTCGAGGCAAGAGGGTGGAATTCCATGAAACAGAGC
AAATGAACGAGGTCAGTATCACTGTGGACTATTCCCAAATTTAGTGTTGCAGTGGAT
ACATTTAGTGGTAACTGCTAATTCTGGTTATATGTGAGTTCATAAGCCAATGGTGTA
GTTCGAATGGGTTAGCCTAATGAGGTAACATAATTATTCTAAAGAAATGTATAGCCT
AAGTAGCCTAATTAGGTTAGTACTAATGATCTTCTTTACCGATCTTCTATTCTTGATA
TTTTTTGTTCAATCCAACAAGTGGCGTGTTTACATGCAGAGAGCGACTGCTAAAGGA
CATGAAGTTGAAGCTGGATGAAGGAAAAAACTACCTGCAACCGATGGCAATAATGC
ATTGCAGGTACTGGGAATTCACTTGAAATTTTAATGTCTTTTACATTCTTCTATTTTTT
GCTGGTGCCACAGTGGTAACAGTGTTGGGCTTTCGAAATATTGCTGGGAGATTATTT
ATTTGTCAATGTACCACATGATAGGGTTGAGGCTCGTCAGGAATGTCAAGTTGCTCC
TGGACGCGGGTATGGTGCCCACAAGGCCGACAGAGCTGAGATCAAGCGCTGAGACC
ATGCCGCCCGTGCCTGCCTGCGGCTACATGTCACACATGGCTAGAGGCATAGGCTAC
CACTACTGTTCTCAGAGGACAACCTGCTCGATAGCTGGCCTAGGACGGCCTTTAGTT
CCAGCCAGGCTTCTCGGCCCGTCTCACTGCTGAACGCTATTGTTGGCGCCTGATTGA
ACGAAGTTGATGCCGGTAGGATCCCATGAAGCAAAGCAACATAAGCAGCAGAAGTC
GGACGAGGAATGATATAGATGCAGACATGCAGTGCCTTCGTTTGTTGCCACTGAAA
AAAATCTGAACCAAGTATTTCTGTAAGTATACCAACAAGTTAGTTGCTACTACTTTG
TATATTGATCTTTGTTGGCTGAATGTTTGAGTTTATTTCAAGGCACAAACACTTTATT
TTGAGATCTGTTCTTCTGTCAGGTTTTACATACTACCTGATTTCAAATGAATTATCTC
ACGCCGATAGAGCGCCCGATAGGGCGCTTTCTATTCTAGTTATAATAGAATAGGTAG
TGCCAAAGCATTGCTAAGACAACACAACTAATATTTTGACCATTTTTCTGCACCTGA
GCAGGCACTAATCGGAGTTATATACAATTAAACTCCTGAAATTTTTTTGACACTTAC
AACTCCATTGACAGTTATGCAGGAACAGTATGACCTACATCTGGACATGACACGGC
GTGATTTTCTTGACTGCACATCAAACTGATGTTCTGGATAGCTGACTGTACTGTATAG
TTGTTTTTTGTGTAGGACATTTGTAGCAAGAACTGAACGTTTTTCTATGTACATATAT
TACATATATATATATATGTCGAAATCTGGTGCTACGAAAACCCAAAACTTAATGAAA
ACATAGTTACCATGCTGGTGAAGATAATGCAGCTAGAGACATATGTTCTAAGTAATA
TCTTCTTTACATAATTTTTCTATGATACCATTTTCTGTTTTGGGTTTTGTCTAGCTATG
TAGATTACAGATTATGATATGCAGTGTCATGTTCTAAATTTGGTTAGTTCTCTTCTCT

GCAATGGTATCCTAGAACTGAATTTTATATCGCGTGGCATGCGACATTCACTAGTAG
AAATATACAAAGCCAATGTCTGGGAGGGAATGCATGCATGTCCAGCTATCCGTGAA
AGAGACTATCTTGTGTGTTGGTACCATGGTTTACATTTTCGTTTTTATTGTTCTGAAA
TTTTGTTTTGATAAAAAAAAATGTTGACTTTCATATTTTAAATTTTGGAACCGAAATGC
GTAGAAAACCGGACTGAAAAATGCATCTTGTTTTTGGAATCCACCGGTTTCTTGGTT
TTCTCGGTTGGTACAACAATTGGTTTCTTCGGTTGGAAAACATAGCACCAGTCGGAC
GGATTTATTAGAAGCAATAGTAGTTTCGATTTATAGAATCCAAATAGGATGTGGATG
ATGATTGTACTACAGATGGACACTTGTTAGTAGTGTCATAGAATGACATCAGGCTTG
TCCCTTATTCTTTTTCACTCATAAATGATAAATCTGCTTCTGCTGTGCTTTCTAGCCAT
TTTAATCGAAACATGCTGGAAGGTTCAATCCCAGACTCCCATCAGCTCGAGCATGCA
CCTCGAACATGTGACTTGAGTCAAAGATTTAACATTTGCATTGTGATTTACTAGGAA
TTTTCTAACCATGATGTCTTTTGTATAGCCTTTTTGATGTAAACAAATTTACTTTGGC
AAATTCCAGCATCTATTGGGGCAATTCCATTACTTACACTTACAGTACTGTAAGTGA
TCATAAAACTTCTATGGTTCAGGAAGCTGCTTATATGTTGTGCTCTCTGAATGAATAC
AGTCTTCTAAACGACAACAGTTTCATGGGACCAATTCCTACTCTCAACAGCGTGACT
AAGCTCAAAGTTTTGTGAGTCGTACAATTACAACATTTAACTTTATTAGGACAACAT
CTGTGTTATGTCAACTGTAATTTTTTTTTTGTGGAAATAACAAAATGCTTTCAAATAA
CAAACTAACTGGACCAATTCCTAATTTGACCAGGATGGGCTTGCTGGAAATTGTGTT
AGAATTTCTTTCAAGCTCGCCAAATTGTGTGTCAGTCGCAGCGCACAGGCATAGTCC
CAGCGGTTCTTTATTTATTTGTTGCAAATGTCTTCTCTCTATGTGCTCGCACCACAAA
ATTCCAAAACCTAGTATCAACCAATGCATTAGCCCTTCTCTTTGTATTCCTCTACTAC
TACTCTGTGATTAACTGATTATACATTTTGTGCCTAGGATTAGCTCGGTACATAACCG
CCATGTCGCAAATTACGACCGCCTGCCACCACTGCATGTTGTAGGTGTTCGAATACG
AGGAGCTCTACGGCGCCATAGCCGACATTAGCTGCTCGTAGAAGATTGTAGAAGGA
ACCTCCCCTCCTCAAAGCACCCCTTCCTCCATCGGGCTGCATCGACGACTTCGCATA
CGAGCCATCTCGGTCAAGCATCAGCCGAGCGGAGCGAGCAAGTCTGCAGGTAAAGT
AAAGATTTGTTTGTTTACATGCAAGCAGTCAATAACCGAGGACACATGATCAGTTTG
TTGTGTTCTGCACTGTCGGTACCCTAAATCAGGGGTACCCTCTTCTACAGCATGAAG
ACGTCGCACCTATGCGACGTCTCCAGGCCACGCGGAGGACGGTGTCTGGCCCCGCC
GCATGGGCAGGCCAAAGGGCGCCACGTGGCAAGAAAAGACAGCACATCCCAGTAT
GTATCGTGGGGTCCGGACCTCCACAGAAGGGCACCAGACCCCTATACGTATAAGTC
CGGAGCGTCCGAGTAAGGTCCGAACCTTGACAAGGTCTCGGAACGAGGAGGACCCT
GGCGTGAGTAGGGGTCCGATGCTGACACGCGTCCAGGCCTTGACTTCCCGCTCAGGC
GGAGACCCGCTGCTTGTGGCCCATGACATAAGCCATCAGACCGAGCCTGACATAAA
GTTGTGCAGCCCCTGCATTTATTGAGGAGGGGACGCGTCGCCTGCCACTGCGCTGAC
GGGCGACGTGCCCTCTCAGCATTTAATGCGTCCCGTCCCATCCGCTGGCAGACAACG
TCAAGGTCCTCCAGCAGGCGCCGCGCCTGCTGGGTCTGCTGGCAAACAGTGCGCCC
GTGCCAAGCAGCGCACTGTGTTATCCCTTCTACAAGAAGCTTCCCCTGCACGCCGAA
GATACACAGACCTCGGACGATAGGACACAAGGAGATTGCCCCGACAGCAAATATTT
ATAGTCCCAAGTGTTATATTCTCTATGTCCCTGGGCCCACATGTCGGGGCTCAGTGC
CTTTGTGCATGCCCTCCTTTAGCTATAAAAGGGGAGGCATGCGACGTTACAAGACAG
ATCCAATCTTAAGCTCAGACACACAATACATTACACAGTGGAGTAGGGTGTTATGCC
CTGGCGGTCCGAACCACTCTAAACCCTCGTGTGCTCTCGTGTGTTCATCCACCCACCT
CGTAACAAGCAAGACGCTTAGGCCCCTCCTCATTTTAGGAATTAGGGCGGGTGCAAT
CCGCCACCCGGCCGGAGAGATTTACCCTCCGACATTTGGCGCGCCAGGTAGGGGCC
TAGGCTTGAAGTTTTTGCTTGTTCCCTCGCTCAGCATGATGGTGCGAATCATCGAGC

ACCACGCTAAAACTTCGGAAGATTTGGCGATGGAGCAAGAGGTTGCGTCCTCTGCG
CCACGAGTACCCGACCGCCCTGAGTCTGGGGCTGCTCCTATACACATCGTGCAACAG
CACACGCCTGCGCAAGTGTCTCGGACTCCGTCGAGGGCAACACCTACGACATTATCT
GCCGCCAGGGAGTTGCTACGCCACCCCCCTAGCTCCACGGCTTCACCAAGGACCATG
AAGCAGTGGCGTGACGACATTGACCGGCTACTCGGTATGGCACATTCTACCTCGACC
AGATCCAGGCCACGGTCATCCCGACGCCAACGCGAAGCCACGGCGTCTGTATGCTC
GCCTTCCGTAAGGGGTGCACAGATTGAAGACCTCCGGGCAGAGCTCAACCGCAGGC
GTACGGGAGAGGACGCCCGGATCTCCCTGGAAAGGGCGCGCGAGCGCCGCCAAAA
CATCGAGGGCCGCAACCTCGATCAAGACTTTGCTGCGGCAACACCGCAGACACCAA
TGGGCACCCGGTCCCAAACGAGTGTCCCCTTGGCCGGCGTGGGCTGCGCCGCCCTCG
CAGACCATCTTTGCGCGGCTTCATGGCCGCCCAAGTTTCGGCCGCACCTGCCGGAAA
AATACGATGGGACATCAAACCCGTTGGAGTTCCTACAGGTGTACGTCACCGCCATCA
CGGCAGCAGGTGGAAACATCGCTGTGATGGCAACGTATTTTCATGTCGCTTTGTCTG
GACCTGCCCAGACTTGGCTCATGAACCTCGCCCCAGGATCCATCTACACCTGGGAAG
AGCTCTGCGCGCGGTTCGTCGCGAACTTCGCCAGCGCTTACCAACAACACGGTGTGG
AGGCCCACCTCCATGCGGTGAGGCAGGAGGCCGTGGAGACTGTCCGGAAGTTCATC
TCACGCTTCACCAAAGTGCGAGGTACTATACCTCGTATCTCTGATGCCTCCATCATC
ACGGCCTTACGACAGGGATACGTGACAAGAAAATGTTGTAGAAGTTGGCCACACAT
GACGTGGAAACTGTCCCCCACGCTCTTTGCTCTGGCCGACAAGTGCGCCAGAGCCAC
CGAGGGCCGTGCATGGCACTCAGCCCCACAAACCGGGGCTGCCCAGCTGGGCGGCT
TGAGTGCCGCCCCCCCGGGATGAGAAGAAGAAAAAGAAAAAGGATCGCGACTACC
AGAAGCCGCGATCCACCGCTCTAGTCGTTGCAGCTACGACCGGGGGCCAGAGCGAC
CACAACAAACACCCACGGCCGCAAAGGGGTAGCGGCGGCTCGTGCCCTGTACACCC
CAACGGTCGCCACAGCGCCGCGGAGTGCCGCGAGATCATTGATCTCGCGAAACGCG
TCAGCGAACGCATCAGCGAGCGGCGTGAGCAGTCTTCCAAGGACGGCTCCCCACCT
CGTCGTCGCCTTGGCAAGGAAAAGGTCGACGACGGCGAGGTAGCCACGGCTGAGCG
AGACCTCGGGTACCAGTCACCCGAGGGGGACCTAAAGGACGTCTTCACTAGAGACT
CCCACTCCGATGGCGACTAGGCGAACCACCGGCGGAACCCCATGTCCCCTGGCCTA
CGGAGTTGAAGCCTGCCTTCCTCCGGAAACCATTCTGGGTCCCCCATGGGTCCAGAC
TTCTGACAGGTTTATCCAGGAGTGCCCAGGGCCAAGGACGGGGGCTCCTTCGTCAAA
TGCAGATGGAAACCCGGGGTCGGGACCTAATCTTACGGAAGTACCAAACTGAATCA
GGGCTCCGCAACACTCCCCAGCTGGGAAGGACCCTTCAAGGTAACGGGAATACTCT
GACCCGAGGGCGACTGTCTAGCTACGACCGAAGGAGTACCTCCCCCGGTGCTAGAA
CATCTCTGTAAGTTCTATCTATAAAGCAGGTCTGAGGGGTCGAGTTATTTCCCTTTTG
TAACTAGGTAGCGCATACGTGTACGCCAACCCGGTGGGGTCTGTCCCATAAACCCG
GCCTGTTGGACCGCACACATGCATGATAAGTTATAAAGGAAAACTACCCTCTCAGAT
GCGCCTCTATGATAGTATTGTCCTACTGCTCCATGGTCTTACCTTGCAGTCTTCTGGA
TGTTATTTTACCTAACCCACCCAAACAATTCTCGTTCCATCAAACATAATGACATCCG
AATTGGACAACCAGGCTTGCAGGTTAGGACCTCTTCGCAAAAGCAGGAGGGCCCGA
TAGCCTGGGGGGCCGGTTCTAAAGAATGGGAACCCGTTCCATGGGGTGGTCCGAAG
CCTGTGTGGCCGCTTAGCCTAGTCCCGTACCCGAAAGCCTGCACACTCCACCGCTCC
GTGACGGGTACCCTAGTACTTAGAATTATAAGTCCTGTGGGTCCGACAACCTGGGGT
CCGGAACTAGAGAAGAGACACTTGTCTCCTAGAGTGGCCCACCTTTGGTAAGCCTTA
CACTTCACATGAACTCCCACCTTTGGCGAGTTCATTCACATTATTCCCCACAACTTGT
TGAGCGATGAACATATGTGAGCTCACTCTTGCTGTCTCACACCCCCACAGGTCAAGA
ACAGGTACCGCAGGATGAGGCACATGGAGGATGCTGCGATGTGTTCGTGAGAGGTC

TAGGTCGTTGTCTCCCAGTCAACTTTGGGTTGCTGGACCGTTGTCTCCTTATAATGTA
AATATTTATTTATTTTGTACAGAACTCCTATTATATAGTAAAGATGTGACATTCAATC
ATGTGCCATGATTCATCATATGTGTGAGACTTGGTCCCAGCACACCTGGTGATTATTT
CGCGCCTGGGTTTTGGACCCCTAAAACCCGGGTGTGACAGAAGTGGTATCAGAGGA
ATGTTGACTGTAGGACGAAACCTAGATAGAACTGGACAACCTATATTTGCTTACCTT
TGCAACTCTGATCTTTTCTAAACTTTTCTTTAGGTTGCCATCCAAGGGGGTCACCCTC
CTTTTAAAGGCGACTCTCCCCACTTGCGTCCTCAGCCGGCGCGGACTGAGTCTTCTC
CAACACGCTCCAAGGTCCTCCCCCTACGACACGGGGGCTGGGTCCCACGCGTCATGC
AAGCTGGCCCAGGACAGAAGAAGCCAAACCGCCGCGCGCGGAGCGCACAACAGCC
CAGCGGTTACAAGCATTCCTCCACTATCGCCCGGACCAGCGGGTGAAAGGGTGGAC
CGCCATGCAGGCGGCATGCAACCGCACCAAGGGGGCGCACCCTTTCGACTTCGACG
CGTCCAGCATGGAGGCCCAGGCCCACACGTCATGTAACCGGCGCGCCGGTTACTAC
GTTCGAGAAACTGCACCGCCACTCGCGCCAGTACCGCGCCTCCTCGACCGCGGAAC
CGGTGCCGCGACTCGAGGCGACCCTGCGCATGACCCGACAGTGCCAACCGAGCATA
TCGGTCACGGGTCAGTCAGCCACGGGAGAAGGCGCGACGGTCGATACGGCCAGAAA
TGGGCCGACAGTAATGGCAGTGGCAGGTGGGCGGAAGCAGCAGTCAAGTCGTCAGC
AAGCTCACGCCTCCTCCTGGGATAGCAGGAGAACCCTCTCCCACGACGTGAAGACG
ACGCGCCCATGTTCCGTTCCTCGAACGGCTCGCGCACGCACAACGGCTGCCCCGCGA
ACCACTCGTCCCATCGCGTTAACTCTACGGCGGGACATGCGGCACCTTTGGCAGGCA
AAGCAGGCGACGCTTCACCTCCGCCATAATGACCGCGTCAAAAAAAAGGTGCGCCA
CATCATTCGATTTCGTATCCTTTTCCACTTCCTCTTTCTCTCTCTTGCCACAGGGACCG
GGAAATGGGATACTCCGAAAGGGATCCTTCTCCGCGAAGGAAGCGGGCCCCGAGCC
CCCCTACTGATCAGAGGTTCGAAGGCTGGCCCCTCGGAAGGGTTCAACAGCCGCCTC
AGAGCACTCGGGCTCCGCGCCCACTACTGGTGTCAGAGGTTCGAAGGCTGTGCCCTC
GGAAGGGTTCAACGACCGCCTCATGCCACTTGGGCTCCGCGCCCACTACTGATCAGG
GGTTCGTAGGCTGGCCCCCGAAGGGTTCGACAGCCGCCTCAGAGCACGCAGAGCGA
GGGATGACCCTAGGTACGTTCGATACATAACCAAGGCTCGGGCTACGCTCCCGAGG
TACCCTAGGACATTTCCGAGACCGACAGGAACGATTATGTAACGGAATCCCACCAG
AGGGC

>B2contig

AGTAGGGTGTTACGCCCCGGCGGCCCGAACCACTCTAAACCCTCGTGTGCTCTCGTG
TGTTAATCCACCAACCTCGTAACAAGCAAGACGCTTAGGCCCCTTCTCATTTTAGGA
ATTAGGGCGGGTGCAATCCGCCACCCGGCCGGATAGATTTACCCTCCGACATGCAC
GTATATTTTTTTGACAAGTGCTKTATTTCTTCTTTCAGATTACTTGATTATTGTTGCCA
TGTGTGWTSTGATCTGCTAACTAAMTGTGTATGTCAAACTTWAWGTTCTTTTTGAGA
ACACGGATATGAACTATTACAACACTTCCACAATATTGCTTACCTAAACATTCAAAC
CGCCTTATTGCGCARCTTCATGGTAGTCGCTTCGATTCTTTTTTTTCCTCGAATATAAT
TATTWGTTTTAGGACAATCGAGATTMCAAATCAACTGGTTTATTGCGATGGTTGGAG
TTCGAGGGTTTACTTTTAGATTAGGTATTATTGCATCATTGGTTTTTTTGTGTGGAAT
TAGCACTTTYTGATATAGTTCCTTGAGTGGTCGCATTTKGCATATGTYATTTGCCATA
TTAATCWAAAACAAATAGGATAAGGATGTAAAATGACATGTCTGCCTGTAGAACTA
TGTATGGTTTGTGTCCTACTTTGCMAMATTCATGGCATTTTGTGTTCATGATAGAAG
GATTCATGTTTCCACTAAGGAGCCAAAAGTCTGGTGAAGCTACATGTAATTTGCATG
ACACCATAGTAGGAGTTTATTCAAGCTATAAAATAAGCGATCTGAAGCTGCATAGA
AGCACGACACCACAGTTTAGAGGCTGGTTCTTAACAACTGGAAACCGCTGTTTCATT
ATGCCAATACCCTTTGGTTTCATTGCTCTTACCTTAGTTCTAACGTTCCCCAGAACTA

CTCACAACTTTGCTCCATTGTCAGCTCCGAGACTTCAMWTTACAAGCTTGCTTTGGT
TATAATTTATGGTTGTGCGAAATGGAATGAMTACCACCATAGGTTTTTATTTGATGT
CYTGGCTGTKTAATAATGGCTAGGCAAATGACACAATTGCTAACTTGCATTTTTGTA
AGTGYYTATTGGGATGATGATWTGGTGGKGATTCAATATTACCAAGGAACATGCCC
TGGTATYCAACTTAYWTTTTTCATACTGTGATTTWMCCTTTTCCWTYTCTCATTWMT
ATTTTTAAGCYTYCAAATAAGAGATGTTATTAATAATATATTATCATATTTTAATTTT
TCAGTGGTGCTTCATTGSMAGCGGAAGTTCATCMACATTGTTCTCTGCACATCAATT
GCTTGTTGTTAGACACTTCTACGRCASGAAAATTTAATCCATCATTGTGTTWTGTAGC
CAACAAATATTATCACTTKGTATGGATGTTTTCGCCAATATCTGGCACATTTCATTTT
GTTTGTTATTAGGTAAGTTACATCTWGCTTTTCTTTTRAWTGSGTTGGACGACTGTM
CGTATACAACTTGAAGGTTCCTCCATGCTTCCATAGAGACCCTGCCCATRAAAGAAA
GGGGGTTGGTTCAGAAACTTGAACCAGAGTACAAWGTTGGTTTGTGAAGTTTGATA
GCATTTCCTAAACAGTTCGAGCAGAGCGTATACAACTACTTATTYMATCACATGATA
CACAAAGCAGGAATCATGAAAACAAGAGTTGACAATTCTTCGAAAAAGATATGARG
ATGCTAGGAAGTAATTCAAGTCTCGATAGGAGCTTTTGTTCCAACATGACTTGTGAA
TCTAAATAAATGATGATAAGGTCGTAAATGCATACTGAAAGTCGCCTAGAGGGGGG
GTGAATAGGGCGAAACTGAAATTTACAAAATTAATCACAACTACAAACCGGGTTAG
CGTTAGAAATATAATCAAGTCCGTGAGAGAGGGTGCAAAACAAATCACAAGCGAAT
AAAGAGTGTGACACGCGGATTTGTTTTACCGAGGTTCGGTTCTCGCAAACCTACTCC
CCGTTGAGGTGGTCACAAAGACCGGGTCTCTTTCAACCCATTTCCCTCTCTCAAGAC
GGTCCCTCGGACCGAGTGAGCTTCTTCTTCTCAATCAAACGGGAACAAACTTCMCCG
CAAGGACCACCACACAATTGGTGTCTCTTGCCTTGGTTACAATTGAGTTGATCGCAA
GGAGATTGAAAGAAAGCACGATTGCAAAAGACCAAGCGVACAAGAGCGACAAGAT
AACACAGCGAATCTCTCTTTCTCTCAAGTCACTAATCACTAATAATCACTTGCCTTAA
CTTTGGAACTKGGAGAGATTGGAGGCTTTGAATGTGTCTYGTGAATGGTATTGCTAR
GCTCTTGTATTGAATGTTAGAAGGTTGGAATGCTTGGGTGTGTTTTGAATGGAGGTG
GTTGGGGTTGTATTTATAGCCACCAACGCATCTTCCTAGGYCGTTTCTSCATTCTGCC
GTGTCGGCGTGCACTKTCCGCACCAGCCTGGTCCGGACGGTCCTCCGCCGCACTGTA
GRTCAACGGACTGRWACGCCAACTGGYCAGCAGTAACGGTCTKATATCTGCATTAA
ATGCGCGCCGTCAGATTGCAGATTAAAGCCAGTCGCRGACGGTCCGGTMGATTGCA
CTCCGGACGGTCGGGACGTCTTATAANKTCAGTCCCCGAGGACCTRTGCACATTCGG
ACKTTTCGCRAGACCGTCACCTTTCGVGTTTTGCGGTCCTGTCACCTACCGGACGGT
CCCGCGCCTGAGGTCCGGACGGTCCGCGCTGAGGTCCTCGGACGCCGTCCTGTGCT
WTKCCTCGGAATGTCCGTAGTGCAGRCTTGGWTTTTCGCATTGGTCTTTGTCTGCAG
CGGTCATCGCTTTGGTGTTTGYGKACGGTCAGCGCTTGCCTGTTTTCKAAAWAGCTT
TCCTGTTTCCGGAATCAATCTACGGTATTCCTGKACAGTCGWTTAGTACTATAGTWA
TAGATGAACTKWGGCACCTGTAGWCAATATAWTSTAGAGGCAAACTAGTATAGTCC
AAWTATTGTGTTGRGCAATGYAASCACCAAATCGATWGGASATAGBTRTAAGCTAA
TTCTCAATCACATASCATAAGTGGAATTCTCAATCGCTATGAATAATGCTACTAATG
ATTATCTACTCGTATTGACGCAAACTGTACAAATCTCTASTCTATCCGWCCCAATGT
AATCAATAGCTATATAATATAAAYATCMACTTTATATGTATTGCTAATACAWATACT
TAKAATAWCAATTATGTTWTTGACGTAACCAATAGTAGTTTATTCTATAATTCMGTA
GCAACGCACKAWACATAGKCCTAGTATATATAAGGCATGRACTGACTAGCCCTTAT
GTTCGAAAWAGAATAACAGAACCTAACCGATATGTATAAATGTGTAACAATCTAAA
CCAACTAATGGGTATGGAGTCTAACATCABAATTCAASTAATAGTMGCRGGATCTGC
AAACAACAATTCAGACTTGVCAGTAGAGGAAATSMAAAACAGTAACAACATCCATT

CSATCACATTGVAGTACCGGCAAGAATATCAGTGGCAAATCCTTGAACTCTGSCGAA
TAAATACAGAGAAAGAATAGCAGWGCTCATAAAYAGACAAATCGCGACGAYCWGA
GCTWAAGGTRTGSAGTGTCGGTCCGCGAAKGAAHCACACAGTATGCACCCTCGCGC
CGACTCGCTGTCGACGGTGGGTGACTCCGGCGACGACCGCGATGCACGGTSTAGTCC
GCACARCTGCGACCSACGGCTGCGCACGCCTGGACAGAGGGGYAGATTCGYCCCAG
MGCCCGTGGTTTGAGCGCCCCCGCAATGTAGCGATYTSTAATTGTCCKACCCRTKCG
TCAACCTAHTCTCTTWWTGAGCGAGCHGCGCMGCGTCCTAACTSCTAATACCAAAS
CAAAATGACGTGTGCGSGATKACCACAGTGTGTGAWGCCCAGTAGGATGAGCTGTG
CGTCATGCCCCTCGYCGGCACGAGMGTCGCGTGMGDGACGTCGTCGYGAAKCGCC
GWTAAACCTCGCGAGGTAAHTTTTGTTACTGATTCGTGTACGGCAGACACTTAACCT
CTCAAGTAATGACTGGACCCAAATACTTCTCAGATATAACGGTACAAAGCCAATTAA
CTCCATACTGGTGGGCTCTTTCGACCTCCGTTGGCCATTCCAATTCCATCTGAATCGA
GAGGCGCCACAACCCCCGCTAACCCCCCGCTCCTAAGAACCGGCGACCGCGCAGCG
CGCVGAGAGCGCCAGCGCACCGCTBGCGGGACCGGGAGGCAGCCACTCGCGGCGCC
GCGCACCGCCGCCGGAGGTCCCCTTTCCAGTTCCATCCGAATCGGAGAGGCGGACG
CCGCGGATCCACCGCTCACCCCCGCTCCTCAAACCTCCAGTGGATCCACCCGCCATG
CCCGCGCGACGGAGGCGCCCGGAGGTTCTGCCACGACAAGGAGAACCCGGCGGAC
GCGCAGCCCGGAGAGCGCCAGCGCACCGCTGCCAGGCGACCGGGAGGCAGCCACTC
GGCGGGCCGCGCCGACGCCGCCTGAGGCCCCCCTTTCCAGTTCCATCCGATCGAGAG
GCCGAGGCGCCCGGATCCACCGCCTCACCCCGCTCCTCAAASCTCCATGGATCCACC
GCTCACCCCCGCTCCTCAAACCTCCATAGATCCACCCGCCATGCCCGCGCCGACGGA
GCGCCCGGAGGTTCTACACAGACAGAGAAACCCAGCGGCGCGCAGCCCGAGAGCG
CCAGCGCCACCGCTGCCGGAACCGGGAGGCAGCACTTGCGGGCGCGCTGCCGCCGC
CGCCGCCGGAGGTCTGTCGCTCTAYATCCTCTGCTCCCGTCCTGCGATGAAGTGCTT
CAGTTAGGTAGTGTTTGCTTCCAGCATTCTTACTGGTAGGATAGAGCCGCTCCGTTCT
AGGTAATGTTTAGTAACTAGTCACTGGCGCTTCCAGCGCGSCGCATTATAATCATTG
CTAATATTAATTTAGAAATAACACTAGAGAAATACAAACACTGTGTTGTCACATTGT
TACATCTAATGTAGGTTGCTTATCAAAGCATAATAACAAAAGGTGGTCTCTACATAT
ATTACATCTAAGTTGGTCGCACAATGGACTGMCATAGTAGCAAAGGTTTTCATAGTT
GATGGTATGCTATTTATTGAATGGAGCGACCATGGAGTATTGGTAGCAWTGCCCAG
TCATCAAATWGCGAGCTTTRACCAAGGTCCTTGCGTCTACAATAGAGAATTAGAGTT
AAATATTGTAAGATTAAATGCGTCTACAAAATACGATATTACTTTACTGAATATTGT
AATGAAAAAATGTGATTTCGGTACCTATTGGACAAATTGTAAGCATGATAATGTGAT
TKGTATTTGAGCAGTTTTGTTGGTAGTAGTAACGCATTAGAAAAAAAGTATAGTTGT
TTAGTGAGGAGGGACAATAAGATATATAGTTATGTAATAAAAGGAGATCTGGTTTG
CAAAAAACAATCAATAACTGGTGTATACCCGTTATGAGTTTCATATTGCTTTGGTAC
TGCATGTTGTCGTATTTCAATGGTGAGCACAGGTAGAAACAAAGGCGAAGTTATAGT
TCAAAGCATCAATATGACCTTGGTTTATTCTTTTTTCTGCTTGAATATATAATGGGGC
GCAGTTCTCCTGCGCGTTTGAGAAAAAAAAAGAATAATCCACACTATTAAGCAATCAT
ATGTGATACATAAGACATGTTTTTAGCGAATAACCAGGTGGGGGCGCGCTATTAGG
AAAAAATAGTAAAAATGGAGCACACTGTTATGAAAATGAATGTGTAATGAGTTGTA
TCCTTGGTCAATTAGCTTTTTTCCCCCTGTGCAGATATGTATAAGTTGACAAAGCTCA
TGTTTCSAAACCAAATATTAGGATTTACTGTCAATTTGAAAAATACCAAAAACAGTA
AGCAGTCAATGCACGGACGTTTCTGTGCGAAACTACTAAGGTACTATAAATACTAAA
TGAATGTTGTGTTTAGAGGTAATGTAGCCAAAACCTGTGAAATGTTTTTCAATAAAT
GTATCACAGTAATACATATAGATATTAAATGAACGGTTGTGTTTAGAGGTAATGTAC

CAAACTTGTGAAAATGTTTGAGTAAAAGTACCCAAGCAATAGATAGTTTATCGCAA
AGAAGAAGTACATACAGAAATAGTTATTGCAAATAATAAATACGTATAGTAACYAC
ACCACTGAAGAGAAATTGTTAAATTGGGAAGGCACTGTTAGGGATAGGATTGCGTA
GTGAAATGTACCTTAATCAAGCAGCCTTTTTTTCCCTTGAGTAAATATGTTGAAGTTG
ATGAAGCATGTGTTTCCAAAACTAAGACTAGAATTTACTGGCAAATTAAAAGAAGG
AAACCATAAGAAGTGAATGCATGGATAATTCCTATGATGAGATACTATGAACATCA
TGTATACAGAAAATGTAGCTACCACTTGTATTGTTTAGAGTTAATTGTAGCAAGCAT
ATGGTCTATACGCAGCACGGCACAACAAAATCAGTATCACTACACACCAAAAGAAT
GCTTTCTGACGAAATGTAACTAAAGTTCAAGAGGCTGTGGCTGAGGATGGTATAGC
AAACTTAAGCGAGGACACTGACATGTACGGAGCTCTGAGGGTGAGGATGTTATAGC
TAACTTATTTTTGACCGACGCCATTAACAATCTGGAAGCACAGGAAGGCTCCACGCG
AAATGAACGGACGACTGATAGCTGCGAGTATGCGGCGAAGCACAGGTAGGGATGTC
AGCAGGGCGTCGGCCCGGATGGTCCCTGTCGGCGACGGCGCGGTGAGTATGCGGCG
AAGCAGGGCATGATGCTGCGGCGGACGACACCGGTGTAGTGCAAGGGCAGTGGACG
GTGGAGGAGGGGGACGTCTAGAAGCCGCGCGACCGACGTCATCTCCGTGGAGAAGA
TGGGCGGCACGACGGAGACGAAGAAGCCCCTGACTTCTTTCCCATTCTCTTCGATCG
ATGATGTAGAAGTTTAATTTATATTCGTATAGCATATGTACGTTCGTATATACAATTA
ATATAGTAAAATAACATTAGTATGCTTTCTCAGCTTGTTATATAATTCAATACCATAG
GGATCTGCTACACCTGTAACCAAACACAGGATGGAGCAGCTCCATTCTACTTCACTA
TAGAACCAAACTGTTCTAGTTACCAAACATAAGAACATAACCGTTTTATTTAGAATC
TAGAATAGAACCGTTCCATTCAACTAGGCTCTAGAACCAAACGCTACCTTAGAAACC
TTATACCGTCGCCTACGTATCTCCTATTACCATCCTCGTGGATGCGTCGATTCAAACT
CGGCAGAATGCATAGATGCCATTTTTTGGAAGGAGAAAAAAAAGACAAATAGTAGC
TCACAACTCACGTGTTTCTTTTTTCTCGAACGCGCAAGAGACATGCGCATCTTTTCAT
TAAGTAGGAAGAGAACAAGATCCAAAATGGATCAGTATCAGTACAAGAAGACCCCC
ATTACGGTGGTCGAAAACAAAGAAACAAAATACATCCATAAATAAAAGGAAACAA
AAACTACTTTGACCTAAAGACCTGGGATTGGGGCTGTTAAGAGATCTAAGTTTGTAG
CCCCTCCTAACACMCACAGGTCCCTTTCCTCCTCGGTTCTCTTGATGGCTTCATTAGA
GGCAGGGAGAAAGGCCGTCAAAAACACAGGCGTTTCGGTGTTTCCACAATATCCAG
AGGCCGAGACTGATAAGCGAGTTTAGTCCTTTTCTGGCAAATCCATGCAACTGATCA
CTACATCTTTYCCACCAGAGCATGGTATTCTCTTCATAGACATAAGGAGTGAAATTC
TGAGGTTAAMCTTTTGTCAGCAGCTTGAACCAGTAATTTCTTGCAACACACATCTCA
CGAGGAGATGGTCAATGTTTCCCATTTCCTGGTCACAAAGGGGGCCACTGATCAGGG
TGATCAAGCCCCTCCTCTGAAGTCTGTCAGCTGTCCAACATCTCTGTAGCGCCACAA
GCCATCAGGAAAAATTTGCATTTGGGGGGGGGACGCAGGAGTGCCAAACCCTGTCA
CAATGCTCAAAATATGTTGATCCAATAAATAAACTCTCGTAGGCTACTTTGGCAGAG
TATTTGCGCATTATGGGCAAACCTAAAGATGTGTTTGTCCTCTCTATGAGGGTTTATG
TTTACTGTGGTAATAGCATCGCCAAAGATCAAGATAATCAGCTAGGACCCTCCACAG
TGAGAGCTCCTTGGATATCATTTAGCCAACCTTCTATTTGCTAAAGCTGTCTGAACA
GTTCTACTGTTTTGAATTCGCCTTGGCACAACTAGTAGCAATCTGGGAGCAATATGC
CATCACATTTCGTCCATGAATCCATCTATCTTCCCAAAAGAGTGTCGAGTGTCCATTG
CCAATTTCCGTGTGAGTGGCAGCATGAAAGAAGGATTTCACCTTGGAAGGAAAATG
CATAGGGAGTCCTGCCCAAGGTTTATCAGGTTCCACTTTAGTCAGCCACAACCATCT
CATTCTCAAAGCCCAACCAAGTTCTTTAAGGCTGAAAATGCCCAATCCTCCCAACTC
AACTGGCCTGCACACTCTACCCCAAGCCACCAGGCAGTGTCCCCCTCTAACTTCTTT
TTTTCCACGCCACAGAAATCCTTTCCTGATTTTATCAATATCCTGTGAGAACTCTTGG

ATGTAAGTCCACTGGGCCATTGCAAGGTAGATGATGCATCCCTGTCATAACGTATTG
GACTTGTACTTTTCTCCAGCTCTAGTCATCAAGTCAGCTTTCCCACGCCCGGGAGCTG
GTCAGCAATTTGTCGATTATGTATTGAGTTTGATTTCTGGAAAGCTTGCTCAAAGAC
AGGGGCAGGCCAAGGTATTTACAGGGGAAAGAAGAGATCTCACAGGGGAGACACT
CCTGAAGGGTCGCCACTTCCTCTTCCCCACATCTAATTGGGTACACACTGCTTTTTTT
GGATGTTGGTTTTCAGCCGCAGAAGCTTTTCCAAAGATCTCTAAAAATTTTGAGAAC
CAGTTGAATATCAAGATTCTTTTGGATGCAAAAACATGATCACATCATCAGCATATA
AGGACACTCTGTGTTGCAAAGGTCTAGAAGCCAATGGATGCAGCAAGATTTTCTAG
ATCAGCTTTGAAATTAACCATCCTAGCAACGTCCATGACAAGTATGAACAACATAGG
TGAGCAGGGGCTTCCCCTGTCTTAAACGCTTTTTTGAGSCTGAATTTTTCGCCAGCAA
CACCATTAAGTAGTAACTTGTGTGGACGAAGTAGATAAACAGTCCACACAGAACAT
TCATCCGMAAATTCTKGCCCAAACCCCCAGTTTTTGAAGAACTTCGAGGAGGAAAG
GCCAAGATAACTGAGTCAAAAGCTTTAGTAATGTCTAGCTTCAAAAGGATTCTTGCT
TGTCTTTGCTGATGGAGAAGTCTGGCAGTCTGTTGCACAATCATAAAATTGTCTTGG
ATGAATCGGCCTTTGATGAAAGCACTCTGCCTTTGGGAGACCATAGATTGCAGTTTA
TTCCCCAGTCTGTTAGCTAGGATTTTAGTGACTAGCTTGGCCATGCTATGCACTAAAC
TAATAGGCCGAAAGTCCTTGACTTGGTCAGCACCCTCCTTTTTTGGAATAAGGGTGA
TGAGAGCTGAATTTAATTTGTGAAAGTGACCAAACTTCCTATCCCAAATGGCTGCCA
AAACTTGCATCACATTCTCTTTAATGATAGGCCAGCAGGCTTTGTAAAATCTTCCAG
TAAACCCATCAGGTCCAGGGGCTTTGTCAGACGGTAGCTCTTTAATGGCTTCAAAGG
CTTCTTTTTCTGTGATTGGTAGTCSCAGATCCTCCAAATTATGAGAGGGCAGCCCCAG
GTAAGTTAAGAWTTAATAGTGACAGTTTCTATCCAGTCTTGTTCCGCAACWAAGCTT
CTGTAGAAGTTATCCACCAGCATGGCTTTKGTCAGCATGCTCAGTAAATACATTATC
TCCAGAGACCAGTTTTKGCCACAAAGTTCTTCTTTCTTTCTATATCTGGCATGCATAT
GGAATAGTTTKGGTGTTAGCATCACCCYCTGCTCTGAGCGCAGTTGATTCTTGATCT
AGTACGTGCAATGGTTTTTCGCAAGTTGATAAVAAKTTATTGTCTCAACTAGAACAG
GTGGCCCAGGCTCTTGAAGGACWAGCAATGGATCAGTGACATCTTGACACGCTCTT
TCCTGGATGGGGATCCAACAATTTTTGCTGATTTGGGATGCTGTTAGCSAATATTGTG
TTAACTCAGGAGGCGGATATTCACACCTGGATACATACTTSCCTCTGGTCAGTTTTCT
TCTAGGTBCTTGTTACAGGGCTTTATTTATGGGAGCTATTACGTTTGAACCTTGGAAG
AGGCTTTGGAAATCTTGGGCTCCCCCTAAGTGGCAAGTTTTTCTTATGGTTGGGCCAT
TCGGAATAAATGTTGGACTTCAGATAGATTGGAGATGAGAGGTCTGGACCACCCAG
TGTCTTGCCTATTCTGTGATCAAGAGCAGGAAACCAATACAACATATTTTGTGTTCTT
GTGTCTTTGCAAGACAGTTCTGGCATTCCATTCTATTTGGATTGAATGTTGGTAATGC
TCTCTCCGCTCAAGAGATGAGATCTCCTTTGCTGATTGGTGGCGAAAAACAAGTAAG
AAGAATACATAAAAGCAAGAGAAAGGGCATGAACAGCATCATCATTTTAGGGGCCT
GGTGTTTATGGATCCACCGCAACAAGGCTGTCTTTGATGGGGAGACCCCTTCGCTAA
GCTCTATCAGATGCGTCTTTCTTGATGAGATTGGGTGTTGGATTACGGCTGGTGCTA
AGCATCTTCAGAGCTTTGGATTAGTGCTTGCGCGTAACCTTCACAGGGTCTAGGGAG
TGGTGTGTTTTTTTTGTTTTGAGTGAGCCTGCTTAGGGCTTTGTACCTTGGTCCATTTT
GGACGCTTTTTYTTCTTCTCTAATTAAAATGATGGCGGCAGTTCTCCTGCGCTTTCTA
AAAAAAAAAAAGTACGTGCAATGGTGTGCGCTTAAATGAGGCAAGGGCTAAGGAAT
GCTTTTTCAGTCCACATCTCAGCCAATCCTCCAAAGGGACAAAAGCCTGAAATCTTG
GGCTACTTCAAGCCTGTGAAGAATCTCCCTTGCTAAAGCTAACTGTGAGGTCACATG
CCCAATCTTTTTGTCCACTCCATTTCTGCAAGTATCTAGCAGTGGCTTTTAGTCTCCT
ATCAAAGCTTAGAACTGGACAGACTCCCTCATCAATAGAAATCCAAGCATCAATTG

ACCACTTCTTGGAATCCATCTAACTTTGGCCAAAAGGATTCAAAATGGAACCGCCTC
TTGCAAAAACAGTTGCTCTGCAGACCTAGCACAAGGGGGCAGTGATCAGAGTCCTC
CGAGGCAGCCGCTTTGGAGGAGGTTATTGGGAAAAATATCTTCCCAGTCACATGAG
CACAGGACTCTGTCTAATCTAACCAATGTAGGAGACACTTGGTGATTAGACCAGGTA
TATTTGCGCCCTATTAGAGGCAGGTCAATAAGGCTTGTATCATCAATAAATTTCCTG
AAAAATCTTCCCAATCATGGCTCTATTGCAAAAAAGCGTTGTTTTTTGTCATCGCTCT
TTATATATAAGATTGAAGTCCCCTAAAACAACGCAACCCAAGGCCCCTGACACTCTG
CTCTGATCTGTCTGAGTTCTAGCAGGAAGGAAATCTTGTCTTCATTGCTTTGGGGACC
ATAAACATTTGTTAACCACCAATGCTGGCCATTGGAGTTTTTGAACTGCACAGATAC
GCTGAAATTATCCACTCTACTCAACCCTGTGACTTGTAAATGGTTCTTCCAGGCTACA
AGAATTCCTCCGCTGGCACCCAAAGAGGGTAAGTAGACATACTGCTCGAAGTTTACT
CCCAGAATGGCTATCACTTTTAATTCGGGTAATTTCCTGCATCTTGGTCTCTTGTAAA
CAAACAATGTCCACTCCCAGGGAATCAATAAAACTGCGAACAGAACTCTGTCTCGA
AGTAGAATTCAGGCCCCTAACATTCCATATAAATATTTTTTTTGGATCCATAAATTAC
ATGTTCAGAAGAACAACAAAAGAGAGAAGTGGACACTCAGCAGGCAGCAGCACTA
TGCGATGAAGCCTCTGCCCGCAGCTCGTCCGGCACCTTCCACCCAAAAAAGGCTGCG
AGAGCCTCTATGTGTTTGAGGGGAAGTCTTTCAGAGAATATTTTTTCATAATCTTTCT
GGTTTGCTGATCTATCTCTTCCAGTTCCCCCCACCACCTTTGAGAGCCTTCATAACTT
TCTGTGTTGCTCTTGAGGAAAGGTCTTGCTTGCCAATTTCTGCTCCCACTCCAGCAAT
GCGGCGGCTGCGACGAAGGGGCACAGCAGTGGACCACTCAGAGATTATGGGCTTTG
GAGCTGGATTTGCTGGTGCTGGCAGAATACCAGAGACTTCTTTCACAATTTGTTTAA
GAAATACCTCTTTTTTGGCTGTTACAGGGTGCGTGCCTTCCATCTGCAGGGCTGGAA
ATGATATTATTTTTGCCGTTGTTTCTTCAGCAATTTCAAACTAGCGGCTTATTATGCT
TCTTTTTTCTATGATAAACGAGCAGGCTTGGTGCCTCTATGAGCCTTCTTTCTGATGT
TGGCAGCGGCCTGGCTTTGGAAACAATGCTGATGGGAATGTTCCTCAAAAATGCACC
GGAGCCGAGGCCTCTCCTAAGGGAGCCTCGCAGGCCCGCACGGTGTTTGGTAGCTT
GGAACCGAGCACAATTCTCTAGTAATACTCTGCTGCCCACTAGGATTGCCAGTTCGA
CTGGTCAGAAAAAACAGCATGTTGAGCATGCTCTTCAGTGTGCACCGGAGCCGAGT
CATCCCGTGGGGATTCTCAAAGGCCCGCACGGTGTTCAGTAGTGTGGTCGCTGCATT
GGTTTTCAATGTGCACCGGAGCGCGAAGCCGCTCTCCGAAGGGAAGCCTCAAGGCC
GCGCCCGGTGAATGAGATTTGGAACCAAGCACAAAATCTCTAGTGAAAATTCTGCT
GCCCCAATGGGAATGCAGTCCATTTGATCAGAAAAAACAGCATGTTGGGCTTGCTCT
TCAATGTGCACCGGAGCCGAGTCATCCCCGTGGGGAGTCTCTAGGCCCGCACGGTGT
TCTCCGTTGAGAATCGCAGCAAGGGCCATCCACAGAATTGTCCCAAATCAGAAGGG
ATAATGATTCGTTTTGTTTGGAATCCAAGCATCTGGGACCCAGTCTTTCTTTGGCTGA
ACGACGTTGCTGTCCCAGCTCCTCAAGATTGCTAGCATTGATGCCTTCCCGTGGTCG
CCCACGTGGAGAGTGTGACCGGAACCGCCGGTATGACCGAGGGCCATCAATCATGA
TCACGGGGCGGACGGCGGTGAGGCAGAGACCAAAAGGAGCAGCAGCAGCTGAAGA
AGAAATAGACCATCTGATCTTCACTGTATAACCCAGAGCAGATGACTTCCTTGCGAA
TCTCCAGTAGGTGCCTGCAGAGGCTCCAGTAATCCAGAGCTCCGTGACTCCGGAATC
ATGGATGTATCAATGCACCAGGCAGAACATCTGAAGGGAAACGCAGGTCACTCAGG
CCAGAGTATCTGAAGTGCACCTTAAATAGCCAAGCAAATGGATTTATTAAATGTTCC
ACACCGTAGCCGTCTCCCACACGTGCACTGGATGCCMATCTAAATTCCAGGTTAACC
AGCTGAGGAAGAGTTCTGCAGAAGAGTTGAAAAAATCTAGACACATCGCTTGCAGA
CCACTGTGAAAGTCGCAGCCCGCTGAATGGACCACCTCGTGTCGAGTCCTTTGAGCC
ATGTCTTCAGACGGGAGAACCAGCAGAAAACTAATTTGGTGATAAGCGACGGAGAA

CAAGGGAGTCTTCAGTGATTTCCAATCTTGGTGAGAGCAGGGATGCAACTTCGAATT
CAGAAAGYTGAGTCAGCGGTTGCTGATCACCGAGACAACCACTGGCCAAACGAAGA
TCTTCCTCGCAACAGTGCATGGCGTGGCTCCAACCTWAGCAAGCATGGGGGTAACG
AGTCGCTGAAGAAGGCAGATGGGGCTGCAAACAAATCCTGCTCCCCATCATCTCTG
GGAAGGGGTGGGTGGTTTCCGTCTCTTGCGAGAGTGACGCCGGCGAGAGGCGGCCC
CTAATAGTGACGGCATCAGCACCCCCCAGCAAGGTCAGAAGGCATTTCTGGAGACA
AATACTGGCTGGCCTGAAGTTGCAGCGACCGAAGTCTKCCTACGCCAAACCTTCGCG
AGGCGCTGCCAGGTTGAATAGATCGCTTCGTCTGGCACAAACCACCATAAGACGGA
GGGGAAAGTCTGCTCCAGACAGACTGCGAAGCCTGCGCCAGGAGGTTGCTGGATTT
CGAAGACGGGACCAGACAGAGGCTTTGCCTGGGGCGACAACCACCAAGACGTGAC
GSGACACACCGACTGCTTGGGAACAGCTGTGGAAATCTGAGCACGATCTGGAGCAC
TCCGAGCAGCGGGATGCCTTATGTCCAAGCCCACAGCACTTGAAGCATCTTGTCCTT
GAACGGCAGTGCACGGACCAGGTGATTTGGGGAGAAGCAGTTGAAGCACCTGCCGG
AGAGATTGGTAGGTGCAGGTGCGCTGGAAATTGCAGCCTTCAGGCCGCAGCCGCCT
GGAGGCGCCGGGACTCAACAGTCTTCCATCATCGCAGCGCCGGCGTCCTTCACAGTG
CCCCCCACGACCGACGGGCTTGAAGAGCCGCAGCCCGGAGCTCCCTACACCGCGAC
GGCCCGTGGCTCAGACCTGGGAGTCCGACTGGCGGAGAAAGAGGGASCGGAGCACA
GGGACCACGTCGCGGAAGGAAGGGTCAGCAGGGACGCTCCCAGCGGAGCTGCAGG
ATCCGCCACAGTCCGCGCGCCWGCGATCCTGTTTGGTCCGCCCCACAGTCCTGGAAC
GGAGCTTCGCACAGAGGGGGAGGATGGGAGGGAGCGGCCGCCTGCAGGGGGCTAT
AAATGGGAGGAGGGAGGAGCGGCGGCGCTCCCCTGTAAGGTGGCCGCCCGACCCG
GAGTGGGCGCCGGCGTGGGGCAATTGGGGAGTGGCAAGCGTCGGCAGGGGAGGGG
AGGGGTCGCGCGGGAGGGAGGGGAAGGGACGTACCTGACGGTTGCCATGTTCTAAC
TGCTTTAACTCACGTGTTTCTCTTCATGAGTTCCTCTTTCATGAGTCATGTGAACTGG
GCGTGTTAATTTCTGGCATATGCTCATTCTCTAGTTCAACTCTGGACACTAGGTGATA
GATGCCCTGTTAATGAATAATGCAATCCTTATTTGGTGWTGTTTAATTGAAATATCC
TGAATGTATTGATTATGCTGCTGTCCTTCTAGTCTTTGCAAAATGAAGCCTGATTTAT
TGAAGGAGGTTGAAAGTCTWTAGAATCGAGCTTCAACATGTAAGACAGGACCGATG
ACAGGTTACTGTCTTATATGGGAACATATAAAGAAATGACTGAAAAATCTGTCATTG
AATTAGAAAACGCCATGAAAAAAGCCTCTAACCCTTCAGGTATGACTATGTGTCTAA
ATAATTTGATCGTTATCTAGATAAATGAAACTGGCTAACAAAGTTTTCAATGACATA
GGAAACATTTCGTCCCAGAGAAATATTATAAGAACATTGAWAAGTAAGCTTGCYAA
CTAGTTGGAGAAAGTTAAAGGTGRCGTACATACCTAATTTTTTTTCATTAATCATTAC
ATGTAATGAACTAATGGTGTTGTGTTATATCTCAAGTAAACACAAAGTTTATTTGTA
TGTTTCTKAACGTTATGAAGTGTTCACATTGTGTAGGAAAAATATTGCATAGTATAA
TTAATATCTCGAACGCTTCTCTTAAATGTACAGATGTCTAATATGATGTCCTCTAGAA
ACCATGACTGAGTATGAGGACATGAAGAAATTGCTAGAGAGTGTCCAGTTAAGCCT
TAAAGAGGCTGAGCAAACAATTTTGAATGGAGAGATTTTACGCAAAAAGTTACATA
ACACAATTCTTGTAAGAATTYTTTTAAACTCAAAAATCTAGTATATGTACACATATG
CTATCCGGCTATTCCTGAATTTGTTTTTTTTTGCAGGAGTTGAAGGGAAACATCAGA
GTATTTTGTAGAGTAAGGCCTTTGTTGCCAAATGAATCTGGAGCGGTCTCATATCCA
AGGAATATACAAGACATAGGCCGAGGCATTGAGTTGTTACATAATGGTAGAGTCAT
CTTTACAATTATTTTCGTAGTTTTAGATTTTGATATGTTGTATTTTAGAATATTGATTA
TGATATATTTGCATTTGCTTATTTTTTCAGCACAAGGATATTCATTTACMTTTTGATA
AAGTATTTGATCATTCCGGCAACYGCAGGAACACGTATTCACAGAAATATCTCAACT
CGTGCAAAGCGCTCTAAGATGGCTATAAGGTGCAACATTAGGATGACTATAACTTA

GATTATCTTTGCAAACCTCTAGAAATAATGATTTGTTGGTTTTTTTTTCCAATATCTA
GTTTTCTTAGAATGCAACTAAAAAAATGGTAAATCATTTTAGGTCTGTATATTTGCCT
ATGGGCAAACTGGTTCTGGTAAAACATACACAATGATGGGCAACCCTGAGTTAGAA
GAACAAAAAGGAATGATACCACGATCTCTGGAGCAAATTTTTCAAGCTACCCAGGC
ACTTAACTCACAGGGATGGATATATAAGATACAGGTGCTTTTTTTTGGTAATTTACTT
ATTTAATTGAACTGACCAGTGAACAGGTTTTTTGTAACTGTTATCATGGTGTATTTTA
AACTTTCAGGCCTCAATGTTAGAAATTTATCTGGAAACCATACGTGACTTGCTTGCG
ACGAAACACATTTCTCAAGATGGTGGTTTTTCAAAGCACAATATCAATGTYATGAAC
GCACAGGGTAGTMATTCGTGTCGGGCCTTGACGGTCATTGACGTAACAAGTATACA
GCGAAGTTTCTTCTCTCCTTAGGCGGGCTGCTCAGAGCAGGTTACACGCTTATTTATT
TTCAAAGTTTCAGCATTAAGTATTGGATCGCAGGAGTATATAATAGCAAAATCTTGT
ACTGATCACGAAATGTTTATGTATGTGCCATAATTTTTGGCAGCAGATTGTAGTATA
TAATGTAGGCTATGTAAAGAAGTCACTACATCATATTCCGTTGACTGTCATATCACT
GCCACTGCGTTCTCAACTTCTTTTGGAGTTCTAGATTTTGAGAATTGAGTGGTGTAGG
ACAGTCATCTATCATTTTGATTTCCAGATCAGTTGGGAGAACACAAATGAATGAGGA
ATCATCCAGAAGTCACTGTGTTTTAAACTCTACTACSGATTTTTGGTGTAAACGAGTG
GTATAATTCTAATTCACATGATATCGTTGTTTGGCATGTTGTCTTTTCGTGCTATAAC
TGACCGTTTTTTTCTCTTCAGAGAATGGACGCAACAAGTGCAAGGAGTGCTCAATCT
TATAGATCTAGCAGGCAGTGAGCGCCTGAACAAAAGTGGTGCCACTGGTGACCAGT
TGAAAGAGGCTGTGGTATTTTGTTATCTAGTCCCTTAGCATTGTTTGTGGTTTTTAGC
TTGATATATATAAATAAAATAACGAAGAGGCATGGTTTGTTCCTACTTAGGCCATTA
ATAAAAGCTTATCATGTTTAAGTGATGTCATCTTCTCAATCGCAAAAAAAGATGAAG
CACGTTCCGTTTCGGAAACTCGAAGCTGACCTATCTGCTCCAGGTATTCCTTATCCTC
TTGCATGCCATCTTGTTTTCGGTGAAAGTCAAATTGTAAATGTATCAAACTGATTCTC
CGCTTCTCAACACAGCCATGTCTCGGAGGGGACTCCAAGACGCTGAATGTTTGTGAA
CCTATCCCCTGAAGTGTCTTCGACTGGGGAGTCTCTCTGCTCATTGCGGTTTGCTGCG
AGGGTGAACTCGTGCGAGATCGGGATCCCTCGGCGCCAAACCAGAATGTCGTCTGA
CCGACCAAGTTGAAATAGGACATCAAGTACTCAGTAAGTCATCATTGAACACCAAG
TCACCAAACGCAGGAGGCTATTAGAATCATGTACATGAATTCAAGTATGTTAATAAT
TATCTCACATTGACTGCAATCCACATATTAACGCGAATATCTGGACTAAGTTACAAC
AAATCTGGAATATCATTAATATTTCTTTTGARGTGTTATTCTGACTGGGTCAGGTGGC
GCCCTTTCTGGTGCCCTTATCCAACACAAGGGAAAACAGTAATTGCAGAGGATCTGA
GAAAATTATTCAGCACAATTGATGTTTTAAAAGGTTGAATGCCCCAGCAAAATTGGT
TAGCGGCGGGAGACAGCTAGAGCAACAAATAAAACATAGAGTAGTAAAACAAAAC
ACAGAGAAGAGCGAGACTAATCATAATTCGCATTAAAGTTGGCAACTGTTTCTTGTT
TTCAACACAAGCCACATACATAGAGGTTGTAGCATAGGCCCTCTACAACAGGGAAC
CGACCCAAAATAAGGTGACAAGGTCCGTACGTAGGGCCAAACACCACAACAGGAG
GTATGTACATGATGCTGGTAGGGGTGGTAATGGATCACGATCCAAATGTTTGTCCAC
AATTTGTTAGGGTCCTAAATTAATTTTAGTTCAAAAATGAATAGAAATAGAGCCCGA
TCCTAACCCGATCTGATCCTTAAATTTTAGCGTAAAGTTTAGATTCCATTGTCAACCC
TAGGTACTGTAAGACATACTCTAGCTAATGCAGTCCACATAGTAATGTACATAAACT
TTAACCAAACCACAAGAATGTGAGTATAAAAGTGCCATAGATATTCCCCGTGATGC
GCTTTTCTTCTCGCAAACCCTGCCTAATTACTTCTAGTTAAAAGATACACGGTTTTGC
TGGCTTCTGAGCGAATTCACGCGGATTCGGGTGTGTTGGCGCAGGCGACGCGTGCCG
TGGTGGGGGTCGTGTTTGGCGGGCTTTCGCGCCATCCCGTTGCGCGACGCGCGTCCT
CCGTGGGTCCGGCGCACGCGGATGTTGAGACCTCGCTTTCGACGGACGCGAAATTTC

CAGCCACACACCTTCTTCTCGGCAGTACAGACGTACAGTGCAGGCAATGGTCCGGC
GGCTGTTCTTAGCTGATGGCGAACTGGCGAAGGCCACGATGCGGGACCCGTCTCTGC
TCCATGCCTCCATTACTTTGTGCCGTCCTCCCTCCTTCCACGCTTCCTTCCCGCAAAG
AGCAGCTTCCCTCCTCTCCTTCCCCVCCAAACGCTCACCTCCCTCTTCTCTCTCGCGC
TGAACTCGACCTAGCGTTTCGACTCCCTCGCGGCCCCCGTCGCAGGGTGTCGTCGTC
CCTGCCACAGTCCGTCGGGGTCGGCTTTGTCCTCGAATCTTCGCACGGATTTTAGGA
GCAGCGACATCAAGATGTTACGGACGCGAGGCCATCAACTCCGCCGCCACCAGCGC
CCGATGCCGGCCCCGGCCCGGCGGATCCCTCCCCGGTGGCCCCGGAGGCTACAACC
CTGGGCAGATGTCGGAGGTGTGGAAGGAGATCTTCGCCTCCGTGCCCGTCCAGTTCC
GCTTCTTCGCTTCTTCAACAGGGAAAATTTCCTTTCTATTGTCACCCCCTGATCGTCC
AGTTTCGCTTAAGTTTCCTTGCTATTGTTAGCAGGAGGCTGTGGCCCGACATACGCT
AGAGGATTTCCTTGCGCCAGGGGGTGGTGATGGAGGGGAACATGGCACTGGCCTGT
CAGCATATTTGCCACCTGCAAATGGCTGATTCTGATTTAATCACAAGGAAATAATAT
GGGTTGTCATTATTGTGGATTATGTCTTAATGTCCTGTACATGCTGGAATCAGGTTGA
CCTTGATCCCTATGGGTCACTAGCTGCATTCCTTGATTCCATGGTTCAATGTGTAGCA
GATGGGGGTACGCTAATGTGTGCTCAGCTATAGACATGATAGTGCTTGCTGGTGGRA
AATGCAGAAGTTTGTTTTTCTAAATGAGAGTTTCAGTTTTAGGATTAAATACGAATT
GCTATCTGCTACACATTGTTTTTCCTAAAACCTTTTGCTTCCAGCGTCTGCCTCCACA
ATACCAACTTTGTGCGAAACACCACTCCTACTCTCCTCAATAAGTACCACATCATCG
GCAAAAAGCATACACCACGGTATATCTCCTTGTATGTCCATGTAAAAATTTAGTGTC
TTCAAAGTTGTCCTTTACTGATTGTTTGGACAAGAAAAATGCCAGTATTTTTTAGGA
GAATTCCAAACTATGATCTACGAAAACAAAATAAAAAGGTCTCCAGTTTATGTTCAT
AATTTAACCTTTTTCTAATATATGAAATATGAAACACAACTAAATGCCATATGCAGT
AAGTTCCTCCTGACATCATGTATTATTTGAGCTATAACTACTTTGTTCTTGACGTAGA
TAAAAGAGTTAAGTTGCTTGCAATAAACTCACAACCTGTGTTATTTCAAGTTACAAA
AAAACCATGTTACAGGTGACATATGTATCCAATTCAAAATAATCATAAATTGCAATC
CAGCAAAAGATCCCTGTGTAAACTGACTGTCCTTTTGCAAGCTCTTGATAGCCACAT
AAAGTGTGAACCAGATAACAAACTATTATTCTACATGCAGTTGCATAGGAAGCATG
GTCACCTTTTCGTTAAAGAAGTAAAAGACTAATAGTTCATCCCGCCGGTGAGGTCTC
AGTTTTAGGTTTCCCTTCGAAGTATATCAACCTGCTATTGGGTTAACCTAGGGTGCAT
GCAGACAACCATGTTGCACTGTGGGATTAATCATGTTGTGGTTTATGGTGGCCGATT
TTGGTTCCTCATTCTCGAGCTGGCTTTATCCTGGAGGTCTGCTCCCTGAGAGAAAGA
ACAAAACAAAAGAACCTAAAAGAAATTTCTGTATTTACATCCACCGTATGGGTTGTA
ATATAACTATGTGTTTTTTTGAGTTCATTGGTTTTGAAGATTCTACTATGATATCTAT
ACATAGTAAGTTGAAATGTTGAGACCTTCAGGACATCGGTTGCTACCACCTGAGGCT
TTAAACTAATCTCAATCACTTCTGCTTTTGCATATTGGTTTGGGTGCATGACATAATT
GGTCAAAGCAATGCTCGAGGCAAGAGGGTGGAATTCCATGAAACAGAGCAAATGA
ACGAGGTCAGTATCACTGTGGACTATTCCCAAATTTAGTGTTGCAGTGGATACATTT
AGTGGTAACTGCTAATTCTGGTTATATGTGAGTTTTTTTTCTGGAATACGCAGGAGA
GCTGCGTATCATTGCATTAATAGAAAGAATAGAAAGGGGGATAGAAATCCCGGTAC
AGAACACACACAACACACACACCAGTGGTAGCCACTTAAAACACTAACTAACTAAC
GGGATCACCGACTATCAAGGAGAAACACAACCGCCAACACTGATGCTAAAGCTAGC
AACCTCATCCACTTGGAGGAAAAGGAACAGAGCCTTGCTGCCAGTCAAGGACCATA
GCACGGCCTCTTCCCTGGCCACATCCTAAGGGGCAGTGACCAGGCTGGATACAAAC
CCCGTCAAAAACGCATCTATTCCTATGCTTCCAAATGGTCAAGCACTAGGATCACTA
GAGAATTAAAACCTTTCTCAAAAAAATCAGATGCCCGAGAGCTGCTAGACTTCCACC

ACGACTCAAAGGAATCCTCATGAGTTGGACAGAACGCTGCAGCCACTAAACTAAGG
AACGAAACCAGAACTGCTTGGCAAAACACACATTTGACCAGCGAAATGATTCATGG
TTTCAGGCCTCTTGGATCGCAAGGGGGGCACTGCTCAGGTCGGTCTAAGTCGCTCTT
TTAATAAGCTGTCAGCAGTGCGCACACCTATTATGTTCCACAAGCCATATAAAGAAT
TGCACTTGCCAGGGAGCCCAAGATTTCCACACACGTTCTGCTGGTTCGAACTGAAAC
CGGAACCAAGGAAAACGAGCCTTTATTAAGCGAACTAGTAAGTATATTCCCCTGAA
GCGATGAATTTCATAAGTGGCGATCCCTTATTCGCAGTTGGAGGTCAACTTGCATCA
AGCCACTTCACAGAGTTCCCAAGAATTTCAGCAATGACTCTAACTGAAATTGCCGCA
TTCACATCGGAGATCCACTGAAAATTAGGGAGCAACATCTTTGACTAAACTGTCTGT
TGGCAAGACGGGGGCCACAAACAACCAACACGAGCAGGAGCTATGTCTTTAAATAC
ATTTCCAAATCGCAACCTCTATGCCTTCCAGAAAAGAGTATTAAGCTCCATCACCGA
TCTTAGTCAATAATGGCTGGCAGCGATCATATTTTTAACCACCTACAAACACGAATG
GGGAAGTCCTCAAGGCCTACTGGGGTCTGTTTTCTGCAGCCATGGCCAGCGAGGCTC
TCAGAGCCCAACATAGGTTCTTTAGCTCACTAATCCCAAGACCACCCAGCTCCTTGG
GACGGGTAACCTTAGGCCAAGCCAAGAGGCAATGACCACCGCGGAACTCTTTCCTG
CCCCTCCAAAGGAAGCCCTTCCTCAATTTATCAATAGCTTTGATGGCCCAAACAGGA
AGGTCGAGTGCCAAAGCTGCATAAATCA

Figure 3.2: FASTA sequences of the B5, C4, and B2 Ab10-I genomic BAC contigs containing the full Kin618 gene model. BACs were sequenced with PacBio technology and assembled using programs optimized for long read length (see text).

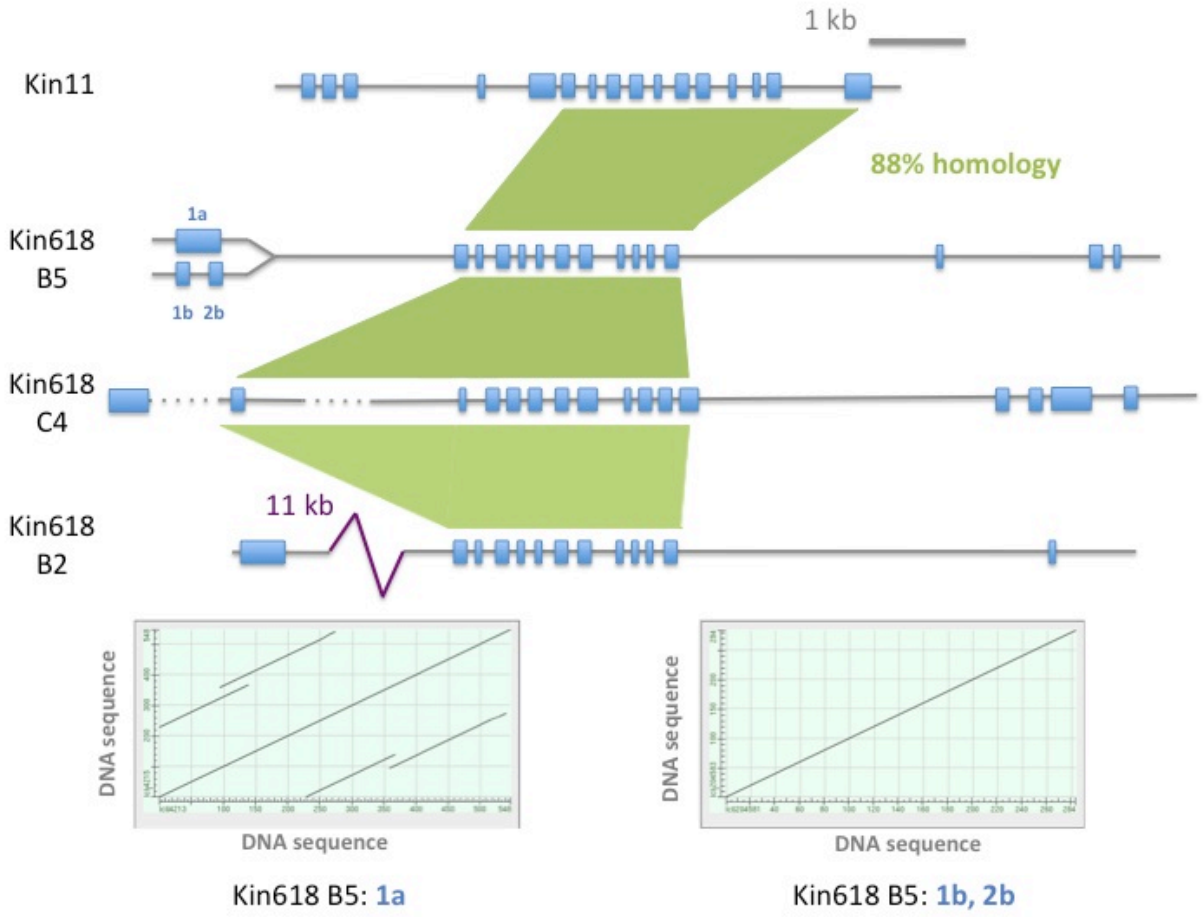

Figure 3.3: Comparison between Kin11 and Kin618 B5 and C4 models as annotated using transcriptome data (see text). Annotation of Kin618 B2 was performed by BLAST alignment of the Kin618 B5 cDNA to the assembled genomic B2 BAC contig. The first intron of the B2 copy spans 11 kilobases and is represented in purple. The C4 copy spans three genomic contigs with dotted lines representing the boundaries between the contigs. Kin11 and Kin618 retain 88% similarity over regions of homology (green). Bottom panels show the sequence of the first exon of Kin618 B5 plotted against itself, both the unspliced (left) and spliced (right) version. The unspliced first exon features a tandem repeat (left; Kin518 B5: 1a).

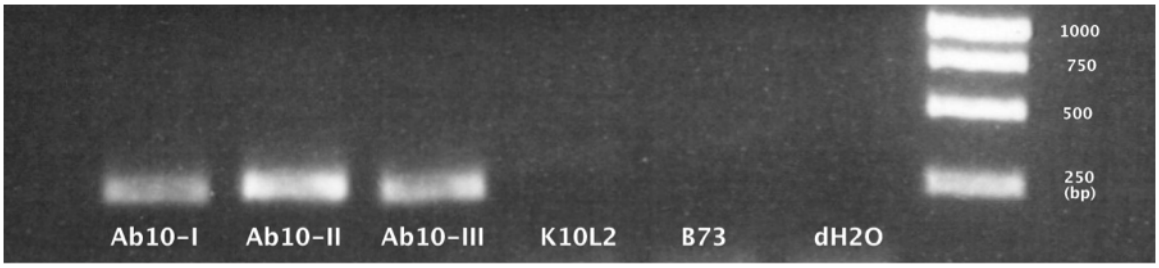

Figure 3.4: Kin618 is specific to the chromosome 10 haplotypes that preferentially segregate. *Kin618 (all)* primers used in assay.

Table 3.1: Summary of high-throughput sequencing data presented in Chapter 3.

Sample	Genotype	Tissue Type	Library Creation	Sequencing
GGF_08_SOL_0001_TGACCA.fq	Ab10/N10	Anther RNA	TruSeq RNA	Illumina HiSeq SE100
A9	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
B2	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
B5	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
B10	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
C4	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
E9	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
F6	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
G11	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE250
B1	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
B6	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
B9	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
C3	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
C10	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
D1	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
D2	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
D6	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
E6	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
E7	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
F2	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
F3	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
F4	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq

				PE300
G2	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
H7	Ab10/Ab10	BAC DNA	TruSeq DNA	Illumina MiSeq PE300
C4_filtered_subreads.fastq	Ab10/Ab10	BAC DNA	5 kb insert	PacBio SMRT
B5_filtered_subreads.fastq	Ab10/Ab10	BAC DNA	5 kb insert	PacBio SMRT
B2PacBio.fastq	Ab10/Ab10	BAC DNA	20 kb insert	PacBio SMRT
E7pacbio.fastq	Ab10/Ab10	BAC DNA	20 kb insert	PacBio SMRT

>Kin618(B5)

CTGCCGGGACCGGGAGGCAGCCACTCGCGGCCGCGCCACCGCCGCCGGAGGTCCCC
CTTTCCAGTTCCATCCGAATCGAGAGGCGGAGGCGCCCGGATCCACCGCTCACCCCC
GCTCCTCAAACCTCCATGGATCCACCCGCCATGCCCGCGCGACGGAGGCGCCCGGA
GGTTCTGCACGACAAGGAGAACCCGGCGGACGCGCAGCCCAGAGAGCGCCAGCGC
ACCGCTGCCGGGACCGGGAGGCAGCCACTCGGGGCCGCGCCGACGCCGCCTGAGGC
CCCCCTTTCCAGTTCCATCCGAATCGAGAGGCCGAGGCGCCCGGATCCACCGCTCAC
CCCCGCTCCTCAAACCTCCATGGATCCACCGCTCACCTCCGCTCCTCAAACCTCCAT
GGATTCACCCGCCATGCCCGCGCGACGGAGGCGCCCGGAGGTTCTACACGACAAGG
AGAACCCAGCGGACGCGCAGCCCGGAGAGCGCCAGCGCACCGCTGCCGGAACCGG
GAGGCAGCCACTTGCGGGCGCGCTGCCGCCGCCGCCGCCGGAGTCTTTGCAAAATG
AAGCTGATTTATTGAAGGAGGTTGAAAGTCTTAAAATCGAGCTTCAACATGTAAGGC
AGGACCGTGACATGTTACTGTCTTATATGGGAACATATAAAGAAATGACTGAAAAA
TCTGTCATTGAATTAGAAAACGCCATGAAAAAAGCCTCTACCCTTCAGGAAACATGT
TCGTCCCAGAGAAATATTATAAGAACATTGGAAAGTAAGCTTGCAACTGTGGAGAA
GTTAAAGATGTCTAATATGATGTCCTCAGAAACCATGACTGAGTATGAGGACATGA
AGAAATTGCTAGAGAGTGTCCAGTTACGCCTTAAAGAGGCTGAGCAAACAATTTTG
AATGGAGAGATTTTACGCAAAAAGTTACATAACACAATTCTTGAGTTGAAGGGAAA
CATCAGAGTATTTTGTAGAGTAAGGCCTTTGTTGCCAAATGAATCTGGAGCGGTCTC
ATATCCAAGGAATATACAAGACATAGGCCGAGGCATTGAGTTGTTACATAATGCAC
AAGGATATTCATTTACTTTTGATAAAGTATTTGATCATTCCGCAACGCAGGAACACG
TATTCACAGAAATATCTCAACTCGTGCAAAGCGCTCTAGATGGCTATAAGGTCTGTA
TATTTGCCTATGGGCAAACTGGTTCTGGTAAAACATACACAATGATGGGCAACCCTG
AGTTAGAAGAACAAAAAGGAATGATACCACGATCTCTGGAGCAAATTTTTCAAGCT
ACCCAGGCACTTAACTCACAGGGATGGATATATAAGATACAGGCCTCAATGTTAGA
AATTTATCTGGAAACCATACGTGACTTGCTTGCGACGAAACACATTTCTCAAGATGG
TGGTTCTTCAAAGCACAATATCAAGCATGACGCACAGGGTAGTATTGTGTCGGGCCT
TACGGTCATTGACGTAACAAGTATAAGCGAAGTTTCTTCTCTCCTTAGGCGGGCTGC
TCAGAGCAGATCAGTTGGGAGAACACAAATGAATGAGGAATCATCCAGAAGTCACT
GTGTTTTAACTCTACGGATTTTTGGTGTAAACGAGAGAATGGACCAACAAGTGCAAG
GAGTGCTCAATCTTATAGATCTAGCAGGCAGTGAGCGCCTGAACAAAAGTGGTGCC
ACTGGTGACCAGTTGAAAGAGGCTGTGGCCATTAATAAAAGCTTATCATGTTTAAGT
GATGTCATCTTCTCAATCGCAAAAAAAGATGAACACGTTCCGTTTCGGAACTCGAAG
CTGACCTATCTGCTCCAGCCATGTCTCGGAGGGGACTCCAAGACGCTGATGTTTGTG
AACCTATCCCCTGAAGTGTCTTCGACTGGGGAGTCTCTCTGCTCATTGCGGTTTGCTG
CGAGGGTGAACTCGTGCGAGATCGGGATCCCTCGGCGCCAAACCAGAATGTCGTCT
GACCGACCAAGTTGAAATAGGACATCAAGTACTCACAATGCTCGAGGCAAGAGGGT
GGAATTCCATGAAACAGAGCAAATGAACGAGAGAGCGACTGCTAAAGGACATGAA
GTTGAAGCTGGATGAAGGAAAAAACTACCTGCAACCGATGGCAATAATGCATTGCA
GGGTTGAGGCTCGTCAGGAA

>Kin618C4

CCCGGAGAGCGCCAGCGCACCGCTGCCGGGACCGGGAGGCAGCCACTCGCGGCCGC
GCCACCGCCGCCGGAGGTCCCCCTTTCCAGTTCCATCCGAATCGAGAGGCGGAGGC
GCCCGGATCCACCGCTCACCCCCGCTCCTCAAACCTCCATGGATCCACCCGCCATGC
CCGCGCGACGGAGGCGCCCGGAGGTTCTGCACGACAAGGAGAACCCGGCGGACGC
GCAGCCCGGAGAGCGCCAGCGCACCGCTGCCGGGACCGGGAGGCAGCCACTCGGG

GCCGCGCCGACGCCGCCTGAGGCCCCCCTTTCCAGTTCCATCCGAATCGNGNGGCCG
AGGCGCCCGGATCCACCGCTCACCCCCGCTCCTCAAACCTCCATGGATCCACCGCTC
ACCCCCGCTCCTCAAACCTCCATAGANCCACCCGCCATGCCCGCGCGACGGAGGCG
CCCGGAGGTTCTACACGACAAGGAGAACCCAGCGGACGCGCAGCCCGAAGAGCGC
CAGCGCACCGCTGCCGGAACCGGGAGGCAGCCACTTGCGGGCGCGCTGCCGCCGCC
GCCGCCGGAGTCTTTGCAAAATGAAGCTGATTTATTGAAGGAGGTTGAAAGTCTTAA
AATCGAGCTTCAACATGTAAGGCAGGACCGTGACAGGTTACTGTCTTATATGGGAAC
ATATAAAGAAATGACTGAAAAATCTGTCATTGAATTAGAAAACGCCATGAAAAAAG
CCTCTACCCTTCAGGAAACATGTTCGTCCCAGAGAAATATTATAAGAACATTGGAAA
GTAAGCTTGCAACTGTGGAGAAGTTAAAGATGCCTAATATGATGTCCTCAGAAACC
ATGACTGAGTATGAGGACATGAAGAAATTGCTAGAGAGTGTCCAGTTACGCCTTAA
AGAGGCTGAGCAAACAATTTTGAATGGAGAGATTTTACGCAAAAAGTTACATAACA
CAATTCTTGAGTTGAAGGGAAACATCAGAGTATTTTGTAGAGTAAGGCCTTTGTTGC
CAAATGAATCTGGAGCGGTCTCATATCCAAGGAATATACAAGACATAGGCCGAGGC
ATTGAGTTGTTACATAATGCACAAGGATATTCATTTACTTTTGATAAAGTATTTGATC
ATTCCGCAACGCAGGAACACGTATTCACAGAAATATCTCAACTCGTGCAAAGCGCT
CTAGATGGCTATAAGGTCTGTATATTTGCCTATGGGCAAACTGGTTCTGGTAAAACA
TACACAATGATGGGCAACCCTGAGTTAGAAGAACAAAAAGGAATGATACCACGATC
TCTGGAGCAAATTTTTCAAGCTACCCAGACACTTAACTCACAGGGATGGATATATAA
GATACAGGCCTCAATGTTAGAAATTTATCTGGAAACCATACGTGACTTGCTTGCGAC
GAAACACATTTCTCAAGATGGTGGTTCTTCAAAGCACAATATCAAGCATGACGCACA
GGGTAGTATTGTGTCGGGCCTTACGGTCATTGACGTAACAAGTATAAGCGAAGTTTC
TTCTCTCCTTAGGCGGGCTGCTCAGAGCAGATCAGTTGGGAGAACACAAATGAATG
AGGAATCATCCAGAAGTCACTGTGTTTTAACTCTACGGATTTTTGGTGTAAACGAGA
GAATGGACCAACAAGTGCAAGGAGTGCTCAATCTTATAGATCTAGCAGGCAGTGAG
CGCCTGAACAAAAGTGGTGCCACTGGTGACCAGTTGAAAGAGGCTGTGGCCATTAA
TAAAAGCTTATCATGTTTAAGTGATGTCATCTTCTCAATCGCAAAAAAAGATGAACA
CGTTCCGTTTCGGAACTCGAAGCTGACCTATCTGCTCCAGCCATGTCTCGGAGGGGA
CTCCAAGACGCTGATGTTTGTGAACCTATCCCCTGAAGTGTCTTCGACTGGGGAGTC
TCTCTGCTCATTGCGGTTTGCTGCGAGGGTGAACTCGTGCGAGATCGGGATCCCTCG
GCGCCAAACCAGAATGTCGTCTGACCGACCAAGTTGAAATAGGACATCAAGTACTC
ACAATGCTCGAGGCAAGAGGGTGGAATTCCATGAAACAGAGCAAATGAACGAGAG
AGCGACTGCTAAAGGACATGAAGTTGAAGCTGGATGAAGGAAAAAACTACCTGCAA
CCGATGGCAATAATGCATTGCAGGGTTGAGGCTCGTCAGGAATGTCAAGTTGCTCCT
GGACGCGGGTATGGTGCCCACAAGGCCGACAGAGCTGAGATCAAGCGCTGAGACCA
TGCCGCCCGTGCCTGCCTGCGGCTACATGTCACACATGGCTAGAGGCATAGGCTACC
ACTACTGTTCTCAGAGGACAACCTGCTCGATAGCTGGCCTAGGACGGCCTTTAGTTC
CAGCCAGGCTTCTCGGCCCGTCTCACTGCTGAACGCTATTGTTGGCGCCTGATTGAA
CGAAGTTGATGCCGGTAGGATCCCATGAAGCAAAGCAACATAAGCAGCAGAAGTCG
GACGAGGAATGATATAGATGCAGACATGCAGTGCCTTCGTTTGTTGCCACTGAAAA
AAATCTGAACCAAGTATTTCTTTATGCAGGAACAGTATGACCTACATCTGGACATGA
CACGGCGTGATTTTCTTGACTGCACATCAAACTGATGTTCTGGATAGCTGACTGTAC
TGTATAGTTGTTTTTTGTGTAGGACATTTGTAGCAAGAACTGAACGTTTTTCTATGTA
CATATATTACATATATATATATATGTCGAAATCTGGTGCTACGAAAACCCAAAACTT
AATGAAAA

Figure 3.5: The transcribed sequences of Kin618 B5 and Kin618 C4 as predicted by alignment of Ab10-I RNA-seq transcriptome reads to the genomic BAC assembly by TopHat (see text). The Kin618 B5 version shows the unspliced first exon.

Table 3.2: Primers used for analyses including genomic primers specific to the B5 and C4 copies of Kin618 and primers used in bisulfite assays.

Product	**Forward Primer**	**Reverse Primer**	**Primer Location**
Kin618 (all)	GCCTTAAAGAGGCTGAGCAA	GCCTCGGCCTATGTCTTGTA	Fourth and fifth exon of Kin618
Kin618 (B5)	GCTACGGTGGGGAATACAAA	CCGGGTTCTTAGGAGTAGGG	Promoter of Kin618, B5 specific
Kin618 (C4)	AAGCCCAGTGATGAGCTTGT	GCCCACCAGTATGGGAGTTA	Promoter of Kin618, C4 specific
B5promoter. Bisulf	CCAAACCCCCRCTAACCCCTACTCC	TGGAGGTTTGAGGAGYGGGGGTGAG	Promoter of Kin618, B5 specific
C4promoter. Bisulf	CCAAACCCCCRCTCCTAAGAACCC	TGGAGGTTTGAGGAGYGGGGGTGAG	Promoter of Kin618, C4 specific
B5intron. Bisulf	ATAACACTARARAAATACAAACACT	GTATTTTGGATGAYTGGGATG	Intron of Kin618, B5 specific
B5MU boundary. Bisulf	GYAAATGAYTGTYTGATGT	ARCCTAARACRCCCAC	Terminal Inverted Repeat (TIR) of Mu and promoter of Kin618 B5

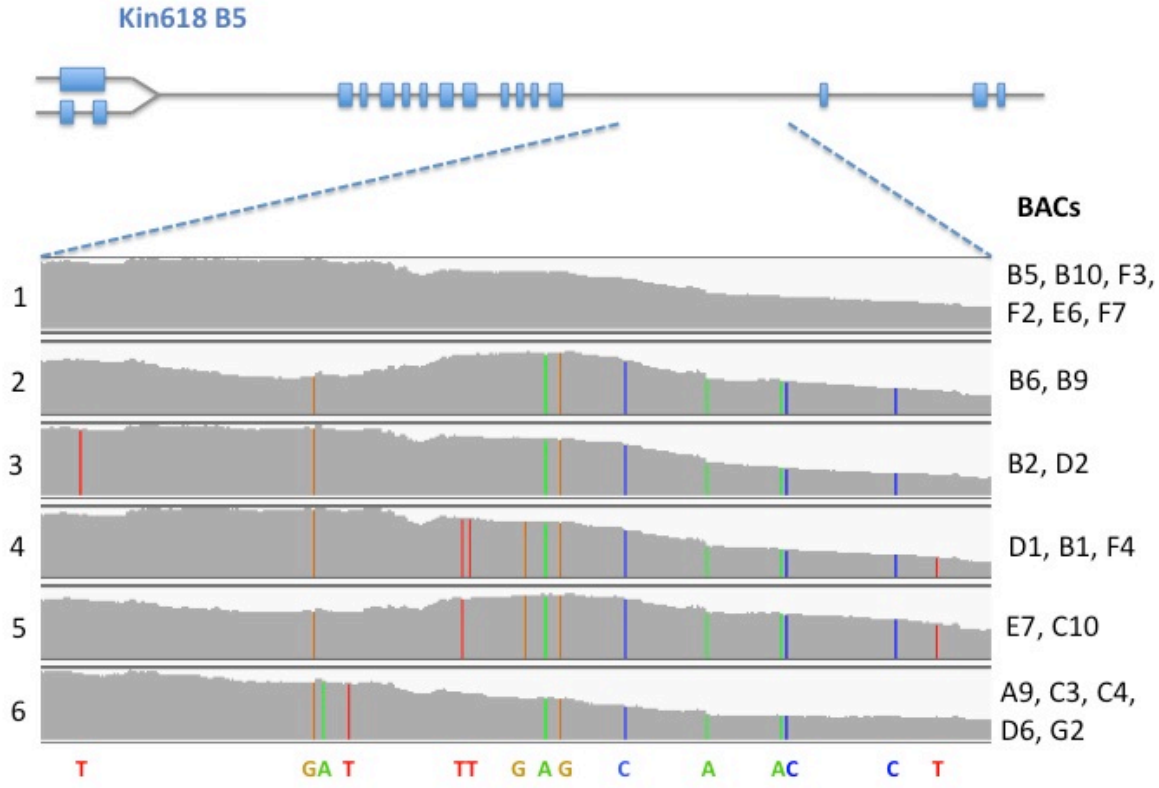

Figure 3.6: SNPs within sets of aligned genomic reads indicate there are six copies of Kin618 within the Ab10 homozygous BAC library. Viewed-in portion features the 12th intron of unspliced Kin618 B5. Genomic reads have been mapped using Bowtie2, with read coverage shown in grey. BACs fall into six categories with members represented by names on right (i.e. group 5 is composed of E7 and C10).

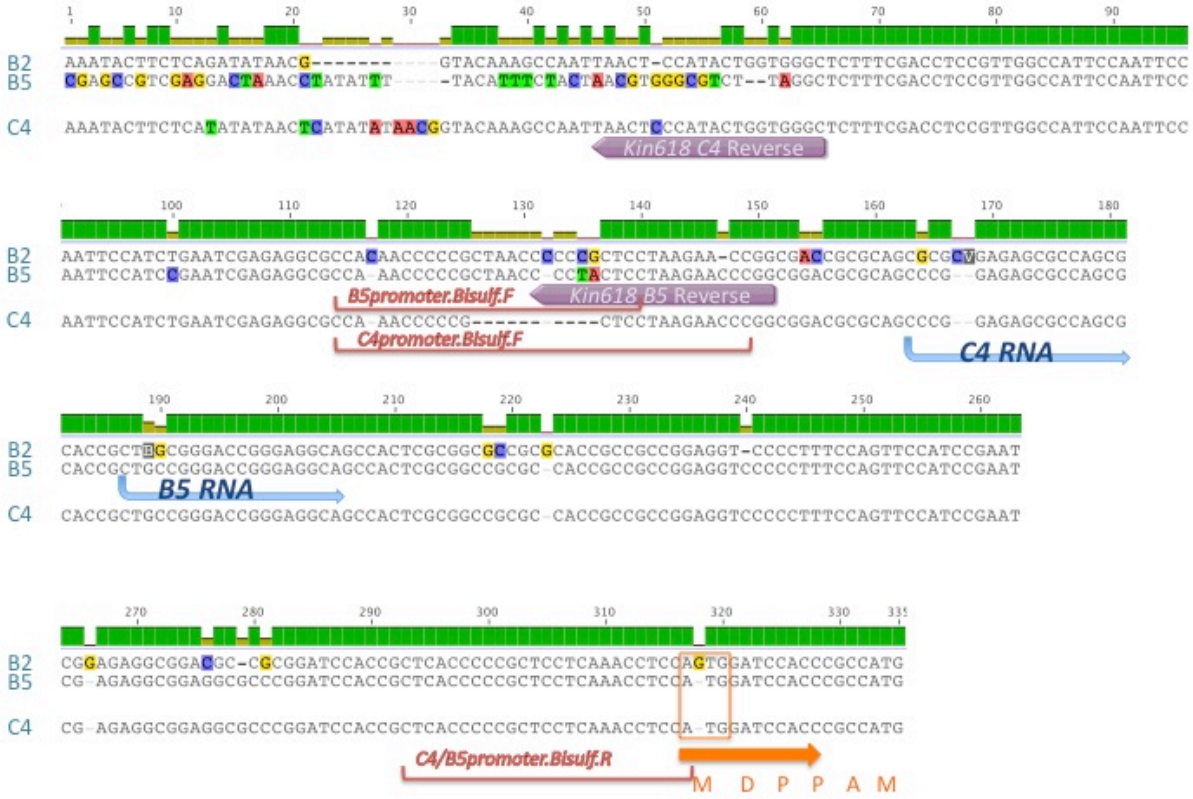

Figure 3.7: Comparison of the B2, B5, and C4 promoter region upstream of the translation start Methionine (orange). Green bar shows level of homology between three sequences. Transcription start cites are shown for C4 and B5 as quantified by TopHat transcriptome read mapping (blue arrows). The reverse gene-specific primers are shown for C4 and B5 (purple arrows). The bisulfite sequencing primers for the B5 and C4 promoters are shown in red. Figure prepared in Geneious.

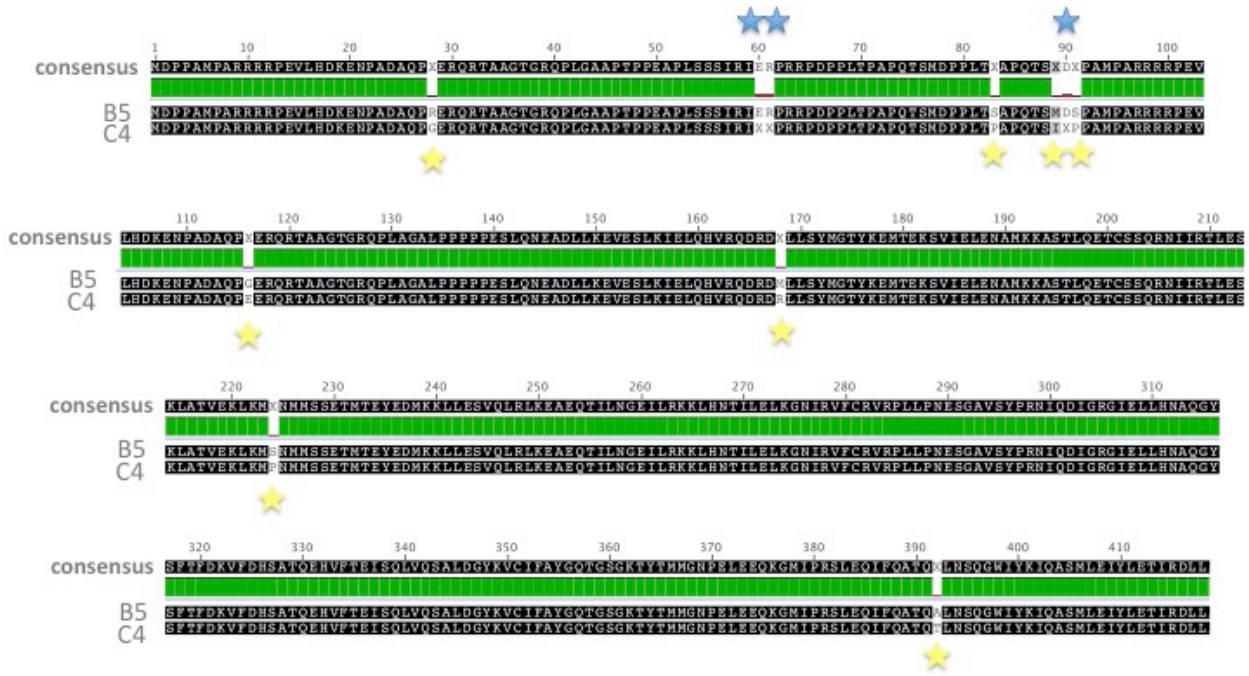

Figure 3.8: Comparison between the first N-terminal 418 amino acids of the predicted protein product of B5 and C4. In this comparison, the unspliced first exon of B5 is used. Green bars represent areas of exact identity. Yellow stars signify amino acid changes. Blue stars represent sequencing errors in the C4 BAC sequencing that prevent comparison (unknown amino acids labeled “X”). Image created in Geneious.

Table 3.3: Lines used to score B5 and C4 diversity/subspecies conservation.

Accession	Subspecies	Collection Site
PI 566692	Zea mays L. subsp. parviglumis	Mexico
Ames 8083	Zea mays subsp. mexicana	Chalco, Mexico
PI 566687	Zea mays L. subsp. parviglumis	Mexico
Ames 21826	Zea mays L. subsp. parviglumis	Mexico
PI 490821	Zea mays subsp. Mays	Brazil
PI 628445	Zea mays subsp. Mays	Jalisco, Mexico

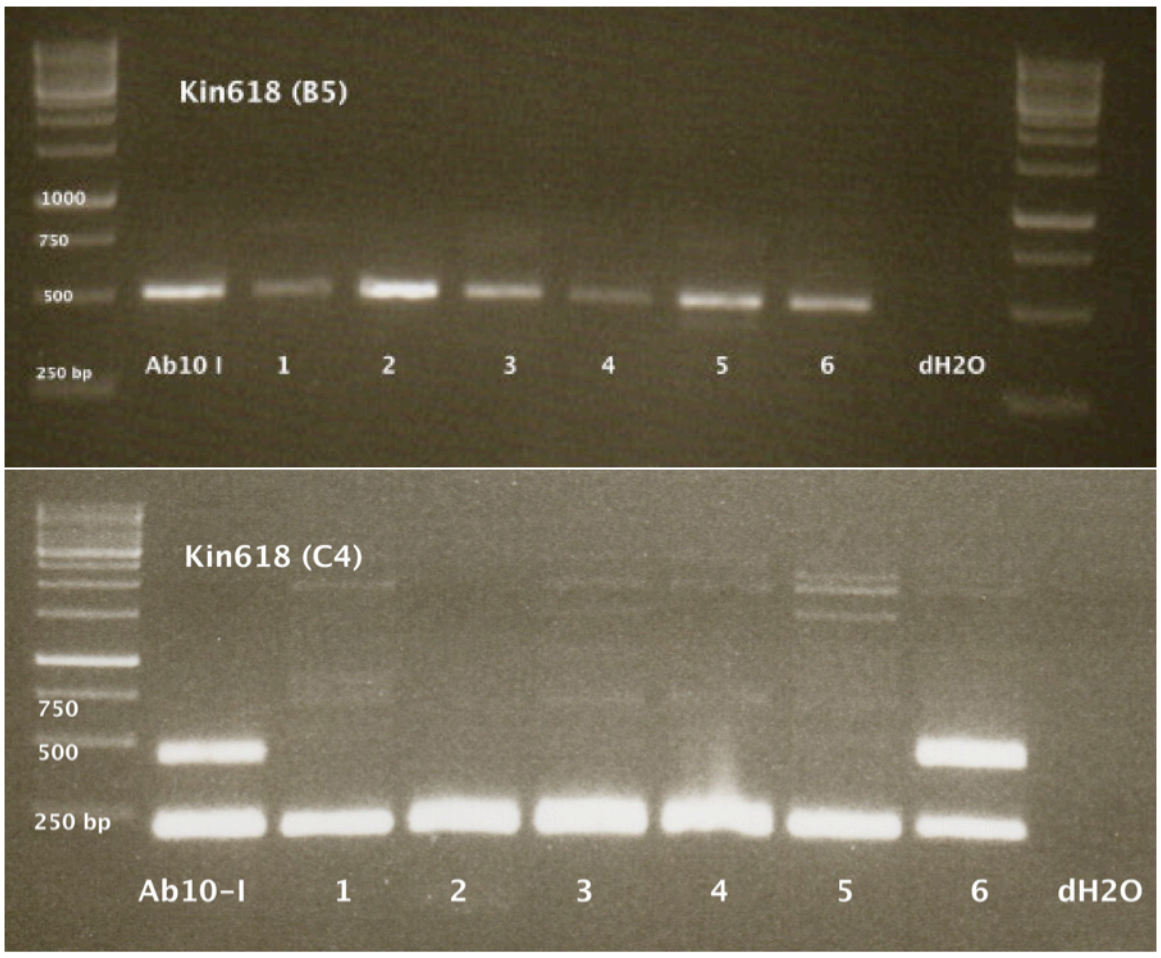

Figure 3.9: Both B5 and C4 are present in the following diverse lines: PI 566692 (1), Ames 8083 (2), PI 566687 (3), Ames 21826 (4), PI 490821 (5), PI 628445 (6). *Kin618 (B5)* primers used to score B5 with an expected product size of 516 base pairs and *Kin618 (C4)* primers used to score C4 with an expected product size of 216 base pairs. The 500 bp band shown in the Kin618 (C4) gel image is a nonspecific product.

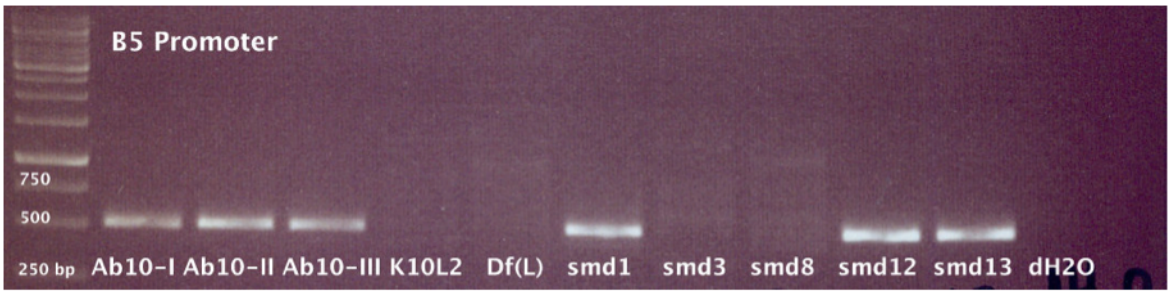

Figure 3.10: The B5 copy of Kin618 is absent in both *smd3/smd3* and *smd8/smd8* using primers *Kin618 (B5)* with an expected product size of 516 bp. Kin618 B5 is also present in three preferentially segregating forms of Ab10 (Ab10-I, Ab10-II, and Ab10-III).

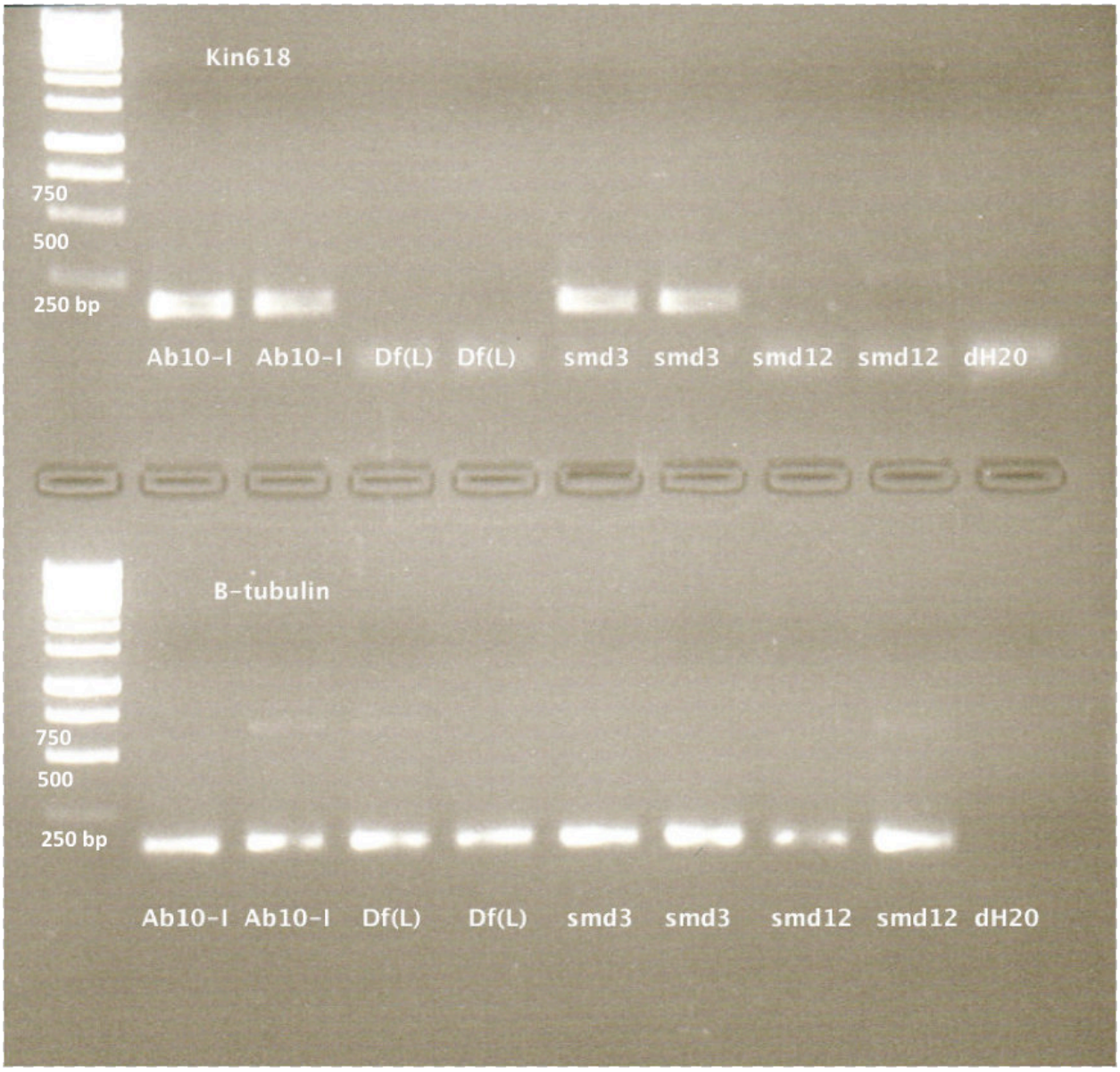

Figure 3.11: PCR on cDNA made from anther RNA indicates that Kin618 is not expressed in *smd12/smd12* as assayed by the primers *Kin618 (all)*.

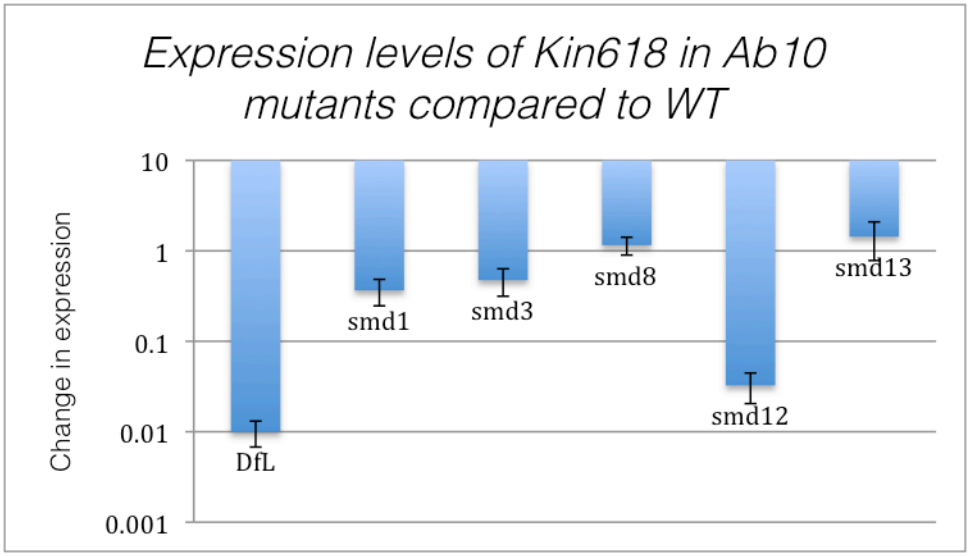

Figure 3.12: *smd12/smd12* and Df(L) shows no expression of Kin618 during meiosis. qRT-PCR was performed by Amy Webster as part of her undergraduate thesis. Change in expression is quantified relative to Ab10-I homozygous plants.

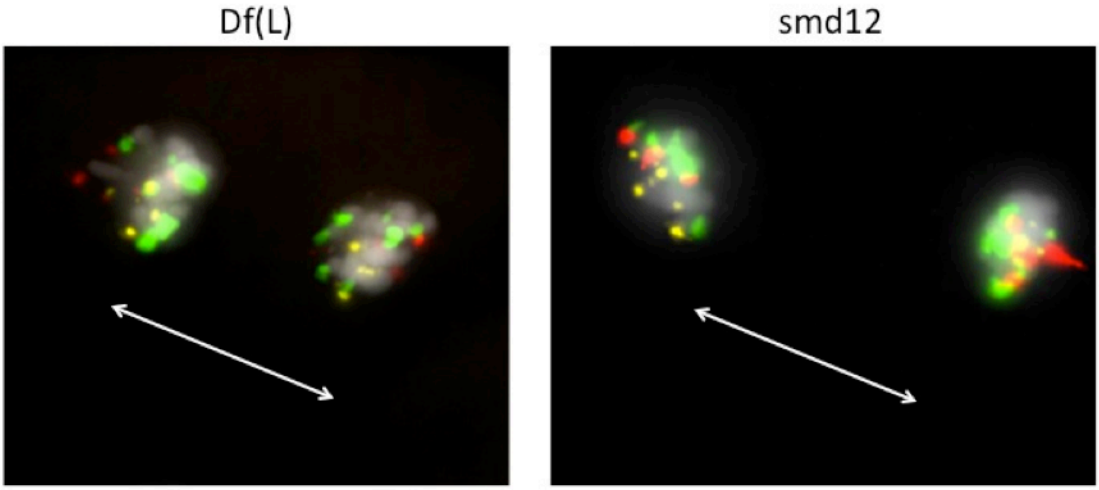

Figure 3.13: FISH image of chromosomes dividing during meiotic anaphase in plants heterozygous for Df(L) and *smd12*. Knob 180 is green, TR-1 is red and Cent-C (centromeres) is yellow. Both Df(L) and smd12 retain TR-1 neocentromere movement but lack Knob 180 movement.

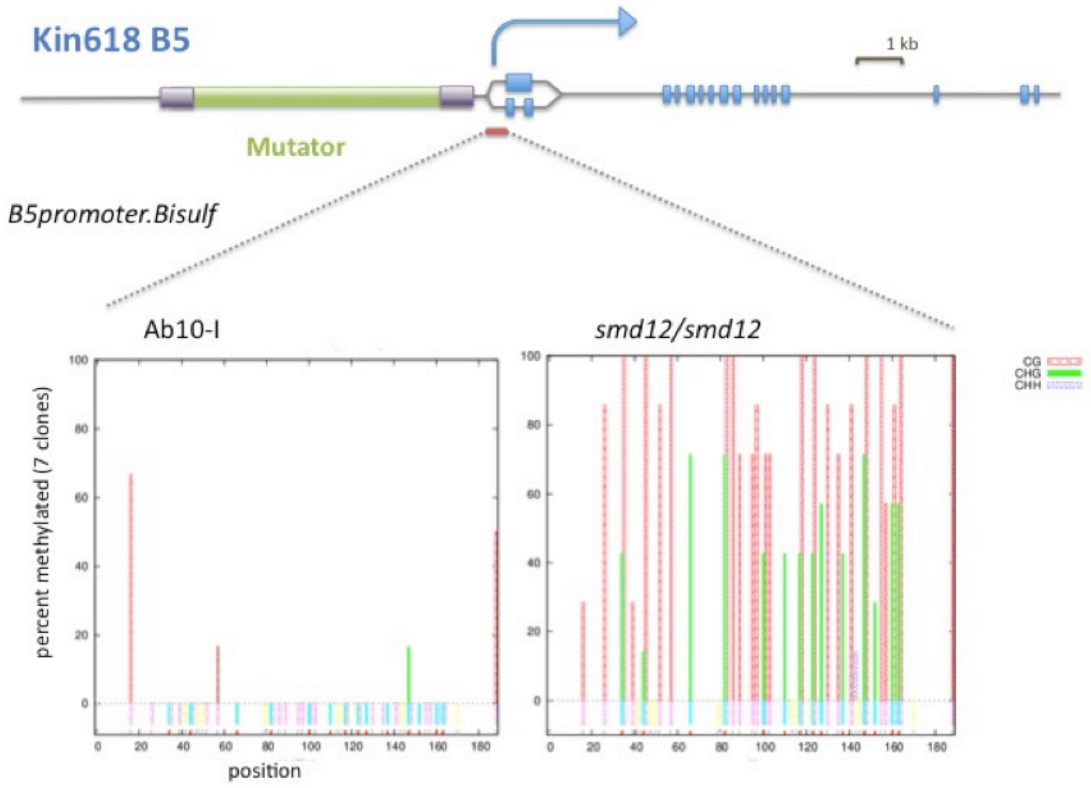

Figure 3.14: Targeted bisulfite sequencing shows an increase in CG and CHG methylation directly upstream of the first exon of B5 Kin618 as measured across 7 samples. Primers used in assay are *B5promoter.Bisulf* and are completely specific to the B5 copy.

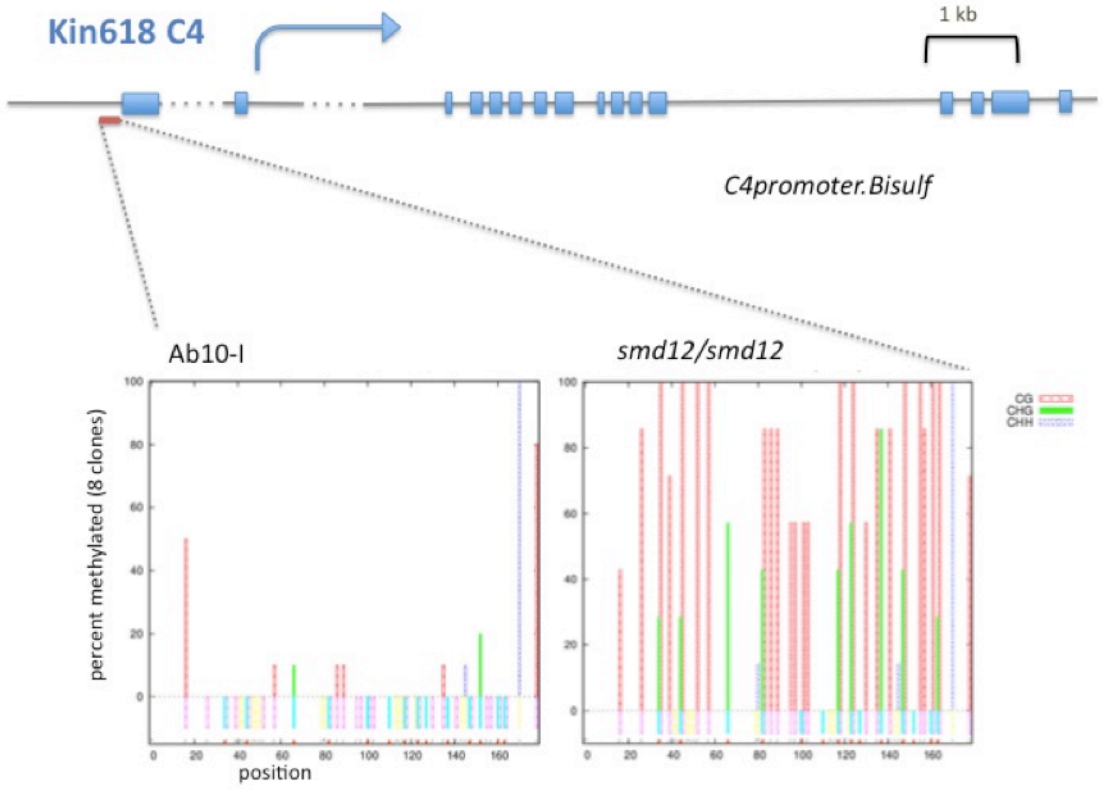

Figure 3.15: Targeted bisulfite sequencing shows an increase in CG and CHG methylation directly upstream of the first exon of Kin618 C4 as measured across 8 samples. Primers used in assay are *C4promoter.Bisulf* and are completely specific to the C4 copy.

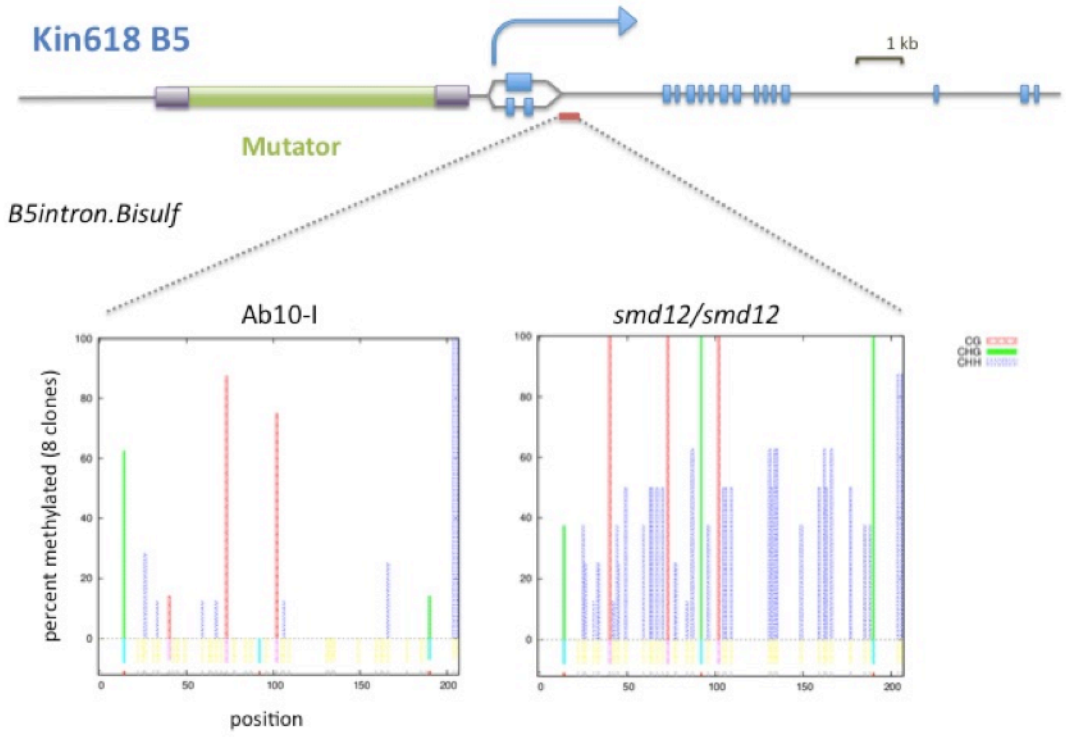

Figure 3.16: Targeted bisulfite sequencing shows an increase in CHH methylation in the first intron of Kin618 B5 as measured across 8 samples. Primers used in assay are *B5intron.Bisulf* and are completely specific to the B5 copy.

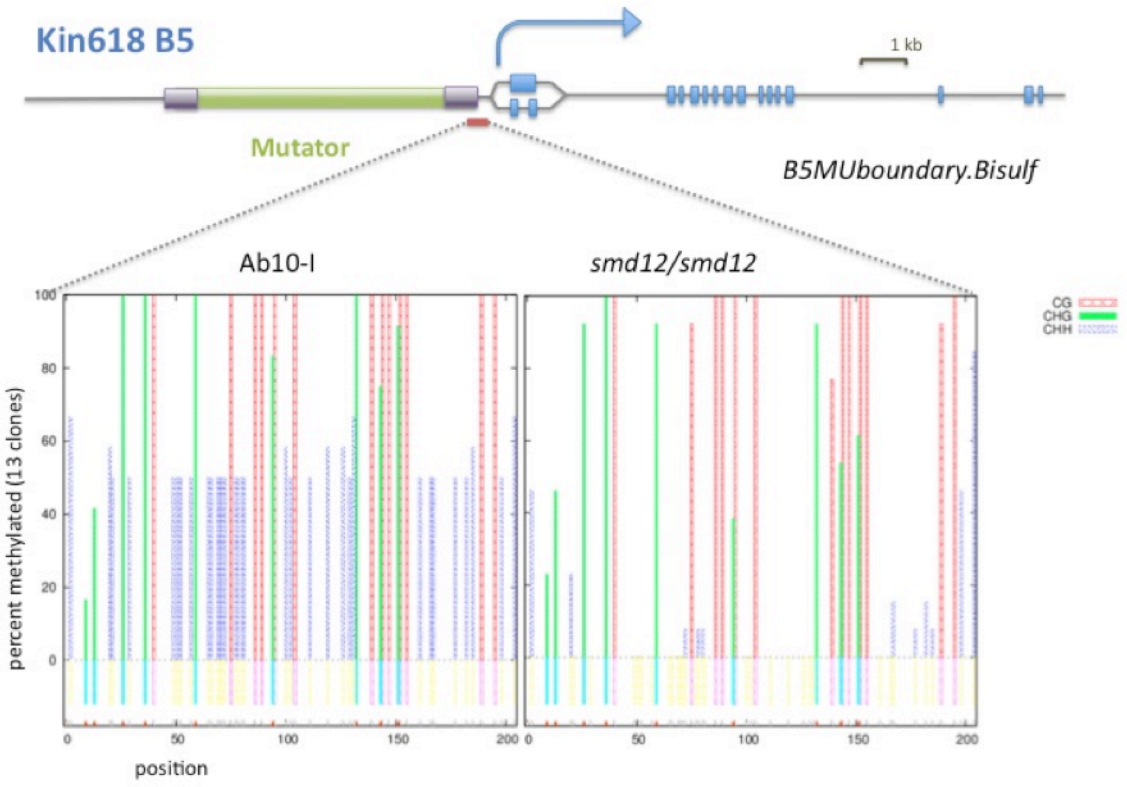

Figure 3.17: Targeted bisulfite sequencing shows a decrease in CHH methylation in the terminal inverted repeat (TIR, purple) directly upstream of Kin618 B5 as measured across 13 samples. Primers used in assay are *B5MUboundary.Bisulf.*

CHAPTER 4: A TRANSCRIPTOME SURVEY TARGETING TR-1 SPECIFIC GENES

Abstract

Abnormal Chromosome 10 (Ab10) in maize is a haplotype variant of normal chromosome 10 (N10) distinguished by its ability to preferentially segregate among progeny at levels over 50%. This meiotic drive function is marked by the formation of "neocentromeres" by two sequence-specific tandem repeats: TR-1 and Knob 180. Only Knob 180 neocentromeres are necessary for preferential segregation and are moved by an Ab10 specific kinesin, Kin618. K10L2 is a variant of N10 that forms only TR-1 neocentromeres and does not meioticially drive. When present in the same maize plant, K10L2 impairs the meiotic drive of other Ab10 types. We hypothesize that a separate, competing kinesin is responsible for the movement of TR-1 knobs. Here we apply a transcriptome approach across maize lines that vary both in geographic origin and the occurrence of TR-1 neocentromere formation to identify the TR-1 gene. We found no transcripts specific or indicative of TR-1 movement, suggesting that K10L2 is more closely related to N10 than Ab10 and that the TR-1 gene may be present yet un-activated in the normal maize genome.

Introduction

Mendel's First Law of Segregation postulates that all genes have a 50% chance of appearing in the offspring, but some genes are able to break this law. We call these genes selfish elements [1]. Selfish elements often carry deleterious alleles that decrease the fitness of the organism. This creates strong evolutionary pressure to select for Mendelian segregation [2,3]. The process by which selfish elements cheat during meiosis is called meiotic drive [4]. Drive can be hard to detect, and may be more common in the natural world than we suspect [5]. Success of the drive system depends on the genetic linkage between a target locus (which is driving) and the distorter of segregation (which is causing the target to drive) [6,7]. This linkage may be broken by recombination: an organismal weapon to thwart the driver [8,9]. When recombination between the target and distorter is blocked, the organism must develop alternative strategies to stop the meiotic drive system. Suppressor elements may evolve to stop selfish genes and restore fairness to meiosis [2,8].

The only true proof that meiotic drive exists comes from the abnormal chromosome 10 (Ab10) system in maize. Ab10 is a selfish chromosome that shows up in the seed between 70-75% of the time, well above the 50% predicted by chance (and Mendel) [10]. Ab10 is a haplotype variant of normal chromosome 10 (N10) that is larger and has knobs. Knobs are large clusters of heterochromatin composed of tandem repeats. Knobs form from two sequences: a 180-bp long sequence (Knob 180) and a 350-bp long sequence (TR-1) [11]. Ab10 has both. Knobs are found on all the other chromosomes too (except N10), however, most maize strains have only a few [12]. Knobs are strange but harmless, except in the presence of Ab10. When Ab10 is present in the plant, all the knobs in the genome meiotically drive.

There are three known versions of Ab10 that differ in knob content but all meiotically drive: Ab10-I, Ab10-II, and Ab10-III [13]. Ab10-I is the best studied and has three TR-1 knobs, a "differential region" containing several genes (L, O, W, Sr) that has been rearranged by multiple inversions relative to N10, a large knob composed of 180-bp repeats and a distal euchromatic tip (Figure 4.1). The inverted segment prevents Ab10-I from recombining with N10 and maintains linkage between distorter and target (knobs). The distal tip of Ab10-I contains the trans-acting factor(s) that cause meiotic drive [14]. When driving, Ab10-I races with both TR-1 and Knob 180 neocentromeres [15]. Interestingly, Ab10-II only activates 180-bp neocentromeres (TR-1 knobs do not form neocentromeres in its presence) but still drives [14]. Ab10-II proves that TR-1 neocentromere movement is not needed to meiotically drive.

Why do TR-1 knobs form neocentromeres if they cannot drive the Abnormal haplotype? A newly characterized Ab10 haplotype, K10L2, only has TR-1 knobs. Maize lines with K10L2 have very fast TR-1 neocentromeres but no Knob 180 neocentromeres. K10L2 has an opposite phenotype to Ab10-II, which only activates Knob 180 but not TR-1. Therefore, Ab10-II only has the genes that activate Knob 180 while K10L2 only has the genes that activate TR-1, making them foils ideal for comparison. Strangely, K10L2 shows extremely weak but statistically significant drive when paired with N10 (51%) [13]. Since K10L2 cannot really drive, its existence eluded explanation until it was crossed into maize plants already containing either Ab10-I or Ab10-II. Fascinatingly, K10L2 reduces the drive of both Ab10-I and Ab10-II from 70-80% to <54% and <60% respectively when they meet within a single plant [13]. K10L2 suppresses the selfishness of the Abnormal chromosome.

Evolutionary theory predicts that genomes favor suppressers of drive [2]. Support for the importance of K10L2 comes from diversity analyses of maize populations. Despite its lack of

strong drive, K10L2 is present in 8% of landrace populations and 42% of teosinte populations, comparable to the presence of Ab10-I (18% and 35%)[12]. These results suggest that K10L2 plays a favorable role in maize populations. This favorable suppression may not be limited just to Ab10-I, but play out across the genome. Other authors postulate that Ab10-I spurs the evolution of knobs, causing them to increase in size and frequency [16]. As knobs contribute upwards of 8% of the maize genome, this can have a serious effect on the genetic content of the plant when knobs start driving [17]. The suppression of Ab10-I by K10L2 indicates that the TR-1 system effectively inhibits the drive of the Knob 180 system. It is possible that TR-1 and Knob 180 are evolving in opposition. As K10L2 and Ab10-I compete for preferential segregation, TR-1 and Knob 180 may also be locked in an intragenomic conflict that alters the genomic landscape of maize.

A previous transcriptome comparison analysis identified the gene responsible for Knob 180 neocentromere movement: a C-terminal kinesin called Kin618 (Chapter 3). Knob 180 and TR-1 move similarly by sliding laterally along the microtubules of the meiotic spindle. We hypothesize that a kinesin similar to Kin618 attaches sequence-specifically to TR-1 knobs and pulls them to the poles of the dividing meiotic cell. Here we perform a transcriptome comparison analysis of meiotic RNA across a series of geographically diverse maize lines, along with one teosinte line, that contain Ab10-I, Ab10-II or K10L2. Two separately isolated versions of Ab10-I, an accession first isolated by Marcus Rhoades and the landrace X233F, provide replicates of abnormal TR-1 activity. A second Rhoades accession contains Ab10-II and does not show TR-1 neocentromeres. A recently identified Ab10-II type in *parviglumis* that has been crossed for two generations into maize. Lastly is a line containing K10L2 with strong TR-1 neocentromeres and the ability to suppress meiotic drive. By examining both the presence/absence and differential

expression of transcripts associated with TR-1 movement, with a special emphasis on kinesin proteins, we thoroughly test our hypothesis that the TR-1 gene is an abnormal-specific kinesin.

Materials and Methods

Genotyping of Ab10 haplotype variants

To isolate the TR-1 motor, two lines containing Ab10-I, two lines containing Ab10-II, and one line containing K10L2, as well as sibling plants without the abnormal chromosomes, were planted in a greenhouse setting (Table 4.1). For nine of the ten conditions, 6 plants were cultivated each. For K10-L2, 10 seeds were planted, for a total of 64 plants. Three weeks later, the entire planting scheme was repeated with 64 more plants. The intent of planting 128 plants was to ensure that there were 32 high-quality anther RNA preps in the end, as shown Table 4.1. DNA was extracted from 2 week old seedlings by the CTAB method described in Chapter 2. The plants continued to grow so that we could later harvest meiocytes.

The presence of the abnormal chromosome is assayed by the genetic marker R which turns the pericarp purple in Rhoades Ab10-I, Rhoades Ab10-II, and K10L2 N10 plants but, to double check that the purple kernels had the abnormal chromosome, we performed PCR with a second marker. Furthermore, the X233F lines and *parviglumis* Ab10-II lines are not marked by R, so approximately 20 seeds were started in a growth chamber and genotyped by PCR before 6 plants were transplanted to the greenhouse for plants of those two lines. Ab10-I and Ab10-II were genotyped with the primers *Kin618 (all)* (Chapter 3), which target the kinesin Kin618 that is unique and specific to the abnormal haplotype. The presence of K10L2 was marked using the primers C2RJJMR which have been shown to cosegregate [18]. Reaction conditions were as

follows: per reaction we used 2X Phusion Master Mix, 0.25 μM primers, 3% DMSO, and 10-30 ng DNA. Reactions were denatured at 98 °C for 5 minutes, followed by 35 cycles of 98 °C for 10 seconds, 55 °C-62 °C for 30 seconds, and 72 °C for 30 seconds, and a final extension at 72°C for 5 minutes. PCR reactions were run on 2% agarose gels.

RNA Extraction and Illumina Sequencing

For each separate plant, the anthers, containing meiocytes, were extracted into PBS and a subset dissected to verify that the anthers were undergoing meiosis and not in the tetrad stage or later. Properly staged anthers were then collected into 1.5 mL eppendorf tubes and flash frozen with liquid nitrogen. Samples were stored at -80°C until RNA extraction. Frozen anthers were hand ground with a plastic pestle in Buffer RF, the stabilizing extraction buffer from the Qiagen RNeasy kit. The protocol from the kit was then followed as specified. RNA quality and abundance was checked on a 2% agarose gel and with the Nanodrop spectrophotometer.

Anther RNA collected from three separate plants was chosen for each genetic condition except for K10L2, where RNA from five separate plants was prepared to allow for biological replicates (Table 4.1). All 32 RNA samples were sent to Georgia Genomics Facility (GGF) for library creation and sequencing. For each sample, a NGS stranded library was created using the Kapa Biosystems kit. All 32 samples were then barcoded and run on Illumina Nextseq using PE75 reads which, given the size of the maize transcriptome, resulted in about 20X coverage per sample.

De Novo Transcriptome Assembly

Quality of samples was checked using FastQC and adapters were trimmed using Trimmomatic (Table 4.2) [19]. A de novo transcriptome assembly was performed by Trinity using 17 of the 32 samples: Rhoades Ab10-I (3), X233F Ab10-I (3), Rhoades Ab10-II (3), *parviglumis* Ab10-II (3), and K10L2 (5) [20]. Trinity subsequently performed alignment of all 32 samples to the de novo assembly using Bowtie and an abundance estimation of FPKM values using RSEM [21].

Presence/Absence Quantification of Transcripts Correlating with TR-1 Neocentromere Activity

FPKM values of all 32 samples were compiled into one file and split into two groups: those containing TR-1 movement (Group A), and those without TR-1 movement (Group B) (Table 4.1). Perl scripts written by Alex Harkess identified contigs that fell into three categories: 1. contigs where FPKM vales were greater than zero for group A (expressed), 2. contigs that equaled zero for group B (not expressed) and 3. contigs that had both greater than zero values for group A and were equal to zero in group B. The analysis was performed again leaving out the 5 K10L2 samples and the 3 N10L2 samples as there was some concern that the C2RJJMR locus was not closely linked to the L2 haplotype.

Differential Expression Between Haplotypes that Vary in TR-1 Neocentromere Activity

Since aberrational read mapping or areas of homology could result in small FPKM values that were nevertheless greater than zero for group B and throw off the entire analysis, a differential expression analysis was performed looking for the genes with the widest fold change in expression between group A and group B. The analysis was again run on Trinity using both the edgeR and the DESeq programs for differential comparison [22,23]. Results were ordered by

their false discovery rate (FDR). As there was some concern that X233F Ab10 and *parviglumis* Ab10-II plants had not been properly scored for neocentromere activity, the analysis was performed a second time leaving out those six samples. A custom script by Alex Harkess employed BLASTX to annotate all differentially expressed transcripts. FPKM expression values of all 134 contigs annotated as kinesins were then closely examined across all 32 samples.

Alignment of Kin618 cDNA to Transcriptome

Since the TR-1 knobs move in a manner similar to Knob 180, we hypothesize that the gene responsible for forming TR-1 neocentromeres is a kinesin related to Kin618. The Kin618 (B5) cDNA sequence was aligned to the Trinity transcriptome assembly using command line NCBI nucleotide BLAST. FPKM expression patterns of contigs which aligned to Kin618 with an e value of less than 0.001 were subsequently examined to see if they mirrored the pattern of TR-1 neocentromeres (high expression across Group A and low expression across Group B).

Results

A Meiotic Transcriptome of the Ab10-I, Ab10-II and K10-L2 Gene Space

The maize species is notable for the large number of genetic differences among lines. In this study, we aimed to harness that diversity and target expressed transcripts unique to our desired phenotype: TR-1 neocentromere activity. We used the same procedure that was so successful in identifying Kin618 (Chapter 2) and applied it to a much wider range of samples and biological replicates that divided into two groups: those with TR-1 neocentromere activity and those without TR-1 neocentromere activity. By extracting RNA from anthers of the desired

genotypes, we narrowed our transcriptome analysis to meiotic tissue. As the Abnormal haplotypes contain genomic sequence not found in the B73 sequenced genome, we conducted a de novo transcriptome assembly with which to conduct our analysis. For the de novo assembly, we used a subset of our samples: those that only contain variants of the 10th maize chromosome (those with Ab10-I, Ab10-II, and K10L2). The assembly resulted in 218, 792 transcripts, which Trinity assembled into 150, 432 "genes" with an N50 value of 1128. The N50 value is the base pair size of the smallest contig such that 50% of the assembly is composed of contigs larger than that value. As this transcriptome assembly represents a much more comprehensive representation of the abnormal meiotic transcriptome than presented in Chapter 2 with multiple genotypes as well as appropriate biological replicates, it presents a valuable resource in future analysis of the abnormal and variable haplotypes.

No Transcripts are Unique to Haplotypes with TR-1 Neocentromere Activity

Once the assembly of the meiotic transcriptome was complete, we were able to quantify FPKM expression values for each gene across the 32 samples that varied both in whether they contained a haplotype variant of N10 or had TR-1 neocentromere activity. Once FPKM values were established, we divided the samples into two groups. Group A, with TR-1 movement, includes 11 samples from two distinct lines with Ab10-I and a line with K10L2: Rhoades Ab10-I (3), X233F Ab10-I (3), and K10L2 (5). Group B is composed of 21 samples from two distinct lines with Ab10-II as well as the N10 siblings from each genetic condition: Rhoades Ab10-II (3), *parviglumis* Ab10-II (3), Rhoades N10-I (3), X233F N10 (3), K10L2 N10 (3), Rhoades N10-II (3), and *parviglumis* N10 (II). A perl script developed by Alex Harkess identified transcripts that contained non-zero FPKM values for Group A and FPKM values that equal zero for group B,

mirroring the presence/absence of TR-1 movement for the two groups. Unfortunately, 0 contigs met this condition. To cast a wider net, conditions were relaxed to target contigs that contained FPKM values equal to 0 across all 21 samples of Group B and could have any value, including 0, for Group A. This condition pulled out 124 contigs, though none of these were expressed in all members of group A (Table 4.3). Since there was some concern that the C2RJJMR locus was not closely linked to the L2 haplotype and K10L2 or N10L2 may have been mis-genotyped, the analysis was repeated leaving out the 5 K10L2 samples from Group A and the 3 N10L2 samples from group B. This revised analysis found only 2 contigs that contained FPKM values greater than 0 for group A and equal to 0 for group B. Neither of the 2, however, were convincingly expressed in the K10L2 samples when they were added back into the analysis (Table 4.4). A final analysis isolated contigs with a FPKM value equal to 0 for group resulting in 66 genes, however, again none of them were convincingly expressed in K10L2.

A *Differential Expression Analysis also Finds No TR-1 Specific Transcripts*

The presence/absence criterion is a strict measure that does not account for abberational read mapping that would be amplified by the number of genotypes and biological replicates used in this study. With this in mind, we quantified differential expression values between Group A and Group B for qualifying contigs. We used two different estimation programs: edgeR and DESeq. The results from both programs were almost identifical, and those from edgeR are presented below. Contigs were ordered in priority based on a high log fold changes and a low false discovery rate (FDR), or the proportion of false positives among all the detections; contigs with FDR values greater than 0.05 are not statistically significant and likely false positives. The 40 contigs that had approximately over a 10 fold increased expression rate in Group A as

compared to Group B were further scrutinized for their sample-specific expression patterns (Table 4.5). Unfortunately, none of the genes with an over 10 fold differential expression level are unique to the haplotypes with the TR-1 motor. The vast majority of the highly differentially expressed genes are transposable elements that have high expression in a limited number of plants in Group A, though are still expressed in a limited number of plants in Group B.

No Kinesin Expression Profile Correlates with TR-1 movement

As TR-1 neocentromeres move in a manner similar to Knob 180, we hypothesize that a kinesin similar to Kin618 transports them laterally along the microtubules of the meiotic spindle. With this hypothesis in mind, we used BLASTX to annotate all contigs that were analyzed in the differential expression analysis between Group A and Group B. We then pulled out the kinesins that showed significant differential expression between the two groups (P value < 0.05) (Table 4.6). Unfortunately, only two of these had a false discovery rate (FDR) of less than 0.05 and only one of the two had a positive fold change, meaning greater expression in Group A than Group B. When the FPKM profile of c95145_g1 was more closely examined, it was barely expressed across all 32 samples, reflected in the logCPM values. In fact, all 9 contigs had extremely low FPKM values across all 32 samples, reflected in the low logCPM values and high FDR values, and none showed a pattern of correlation with TR-1 neocentromere movement. Going further, we looked at all 134 contigs created by our de novo transcriptome analysis that were tagged as kinesins. Carefully examining FPKM values for each contig across all 34 samples, the results were undeniable. There are no kinesins that are differentially expressed, even at a low level, in a way that correlates with TR-1 neocentromere activation.

As there was some concern neocentromeres had not been properly scored in X233F Ab10-I and *parviglumis* Ab10-II, a DESeq differential expression analysis was performed without these 6 samples and the most differentially expressed annotated kinesins were again identified. However, this abbreviated analysis identified the same contigs as those presented in Table 4.6.

Kin618 is robustly expressed in meiosis

The similarity between TR-1 and Knob 180 neocentromeres also leads us to hypothesize that the TR-1 neocentromere gene shares a common ancestor with Kin618. With this theory in mind, the cDNA sequence of Kin618 B5 was aligned to the transcriptome assembly using BLASTn and contigs that BLASTed to Kin618 with an e value less than 0.001 were further analyzed. This approach yielded 19 candidates (Table 4.7). The best hit is Kin618, which showed the expected expression pattern across all 32 samples: only expression in samples with Ab10-I or Ab10-II. All 19 samples were subsequently aligned to the B73 maize genome by BLAST for annotation purposes. Nine of the contigs were native maize kinesins. Finally, the differential expression profiles of all 19 contigs between Group A TR-1 activity and Group B no TR-1 activity was included. All 19 contigs either are barely expressed with low FPKM values or there is little difference between expression in Group A and Group B, leading the FDR value to be equal to 1. Kin618 shows no differential expression between Group A and Group B as expected since there are 6 samples containing Ab10-I in Group A and 6 samples containing Ab10-II in Group B, all of which have comparable FPKM values. Contig c58191_g1 does have statistically significant differential expression (P Value < 0.05) however, it has a negative logFC, indicating that there are higher FPKM values in Group B than in Group A. When more closely examined

across all 32 samples, c58191_g1 has an uninspiring expression pattern with most FPKM values close to or equal to 0, even across 12 of the 21 samples in Group B. None of the Kin618 homologues present an expression pattern predicted of a TR-1 motor.

Discussion

The unusual nature of neocentromere movement makes it especially surprising that it occurs in duplicate. Though Knob 180 neocentromeres are clearly associated with meiotic drive, the function of TR-1 neocentromeres is less certain. The drive suppressive ability of K10L2 suggests that TR-1 evolved to compete against Knob 180 in an arms race scenario [13]. The large, heterochromatic Knob 180 tandem repeats offer no benefit to the maize genome, and the TR-1 repeat may proliferate to prevent Knob 180 knobs from growing too large. In fact, *Zea* species with fewer TR-1 repeats have larger Knob 180 loci [24]. Despite this competitive nature, the two tandem repeat types also work in concert. When both repeats are being driven by Ab10-I, TR-1 may even enhance the preferential segregation of Knob 180.

This close association and similar manner of movement leads us to hypothesize that TR-1 is moved by its own highly-expressed meiotic kinesin. Knob 180 and TR-1 have long been known to be controlled by separate trans-acting factors, and are genetically mapped to different physical positions on Ab10-I [15]. The meiotic drive mutant *smd12*, which lacks preferential segregation and Knob 180 neocentromere movement, still has TR-1 movement (Chapter 3). As the gene Kin618 is silenced in *smd12*, it is clear that TR-1 does not coopt the Knob 180 kinesin for its own transfer. The haplotype variant K10L2 also does not have the gene Kin618, yet still presents with strong TR-1 neocentromere movement.

The gene for control of TR-1 has been mapped to the "differential region" on the Ab10-I chromosome, which contains markers that are also present on normal chromosome 10 [14]. A previous phylogenetic dating analysis of the haplotype variants of chromosome 10 using markers that map from that same genetic region indicates that K10L2 is more closely related to normal chromosome 10 than either Ab10-I or Ab10-II [25]. There are likely very few genetic differences that distinguish K10L2 from normal chromosome 10. Perhaps whatever factor is coopted by TR-1 knobs is transcribed by normal chromosome 10, and thus would not be picked up by our presence/absence assay. This "normal kinesin" may gain TR-1 specificity by an additional factor such as a short, non-coding RNA that would escape our transcriptome survey. With the thought that a kinesin moves TR-1 knobs, we closely examined the expression profiles of all kinesin transcripts across Group A with TR-1 neocentromeres and Group B without TR-1 movement, yet it was abundantly clear that none of the kinesins identified in our study are differentially expressed.

Though TR-1 neocentromeres move similarly to Knob 180, there are some crucial differences. TR-1 repeats string out along the microtubules during anaphase, while Knob 180 repeats remain bunched [15]. Perhaps TR-1 movement does not involve canonical motor proteins but rather some other mechanism, also present on normal chromosome 10. This TR-1 mobility factor may be only subtly more expressed in those haplotypes with TR-1 neocentromeres and remains buried in our differential expression analysis results. As we uncover more of the proteins involved in the meiotic drive phenotype, we may gain insight towards potential mechanisms of TR-1 movement. Furthermore, the large amount of data generated by this chapter provides a comprehensive picture of the meiotic transcriptomes of Ab10-I, Ab10-II, and K10L2 and will be highly useful in further characterization of the meiotic drive system in *Zea mays*.

Acknowledgements

As mentioned in the text, we would like to thank Alex Harkess for extensive bioinformatics help and the construction of custom Perl scripts to complete this analysis.

References

1. Dawkins, R. *The Selfish Gene*. (Oxford University Press: 1989).at <http://books.google.com/books?id=WkHO9HI7koEC>
2. Crow, J. F. Why is Mendelian segregation so exact? *BioEssays: news and reviews in molecular, cellular and developmental biology* **13**, 305–12 (1991).
3. Leigh, E. G. How does selection reconcile individual advantage with the good of the group? *Proceedings of the National Academy of Sciences of the United States of America* **74**, 4542–6 (1977).
4. BURT, A., Trivers, R. & Burt, A. *Genes in Conflict: The Biology of Selfish Genetic Elements*. (Harvard University Press: 2009).at <http://books.google.com/books?id=8e8xLbmtuMcC>
5. Lyttle, T. W. Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends in genetics: TIG* **9**, 205–10 (1993).
6. Prout, T., Bundgaard, J. & Bryant, S. Population genetics of modifiers of meiotic drive. I. The solution of a special case and some general implications. *Theoretical population biology* **4**, 446–65 (1973).
7. Charlesworth, B. & Hartl, D. L. Population Dynamics of the Segregation Distorter Polymorphism of DROSOPHILA MELANOGASTER. *Genetics* **89**, 171–92 (1978).
8. Haig, D. & Grafen, A. Genetic scrambling as a defence against meiotic drive. *Journal of theoretical biology* **153**, 531–58 (1991).
9. Brandvain, Y. & Coop, G. Scrambling eggs: meiotic drive and the evolution of female recombination rates. *Genetics* **190**, 709–23 (2012).
10. Rhoades, M. M. Preferential Segregation in Maize. *Genetics* **27**, 395–407 (1942).
11. Ananiev, E. V, Phillips, R. L. & Rines, H. W. A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10785–90 (1998).
12. McClintock, Yamakake T, B. A. Chromosome Constitution of Races of Maize: its significance in the interpretation of relationships between races and varieties in the Americas. *Colegio de Postgraduados: Chapingo, Mexico* (1981).
13. Kanizay, L. B., Albert, P. S., Birchler, J. A. & Dawe, R. K. Intragenomic conflict between the two major knob repeats of maize. *Genetics* **194**, 81–9 (2013).
14. Mroczek, R. J., Melo, J. R., Luce, A. C., Hiatt, E. N. & Dawe, R. K. The maize Ab10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. *Genetics* **174**, 145–54 (2006).
15. Hiatt, E. N., Kentner, E. K. & Dawe, R. K. Independently regulated neocentromere activity of two classes of tandem repeat arrays. *The Plant cell* **14**, 407–20 (2002).
16. Buckler, E. S. *et al.* Meiotic drive of chromosomal knobs reshaped the maize genome. *Genetics* **153**, 415–26 (1999).
17. Dennis, E. S. & Peacock, W. J. Knob heterochromatin homology in maize and its relatives. *Journal of molecular evolution* **20**, 341–50 (1984).
18. Kanizay, L. B. The variants of maize chromosome 10 and their roles in meiotic drive. 111 (2011).

19. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30**, 2114–20 (2014).
20. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**, 644–52 (2011).
21. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**, 323 (2011).
22. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139–40 (2010).
23. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome biology* **11**, R106 (2010).
24. Albert, P. S., Gao, Z., Danilova, T. V & Birchler, J. A. Diversity of chromosomal karyotypes in maize and its relatives. *Cytogenetic and genome research* **129**, 6–16 (2010).
25. Kanizay, L. B. *et al.* Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in Zea mays. *Heredity* **110**, 570–7 (2013).

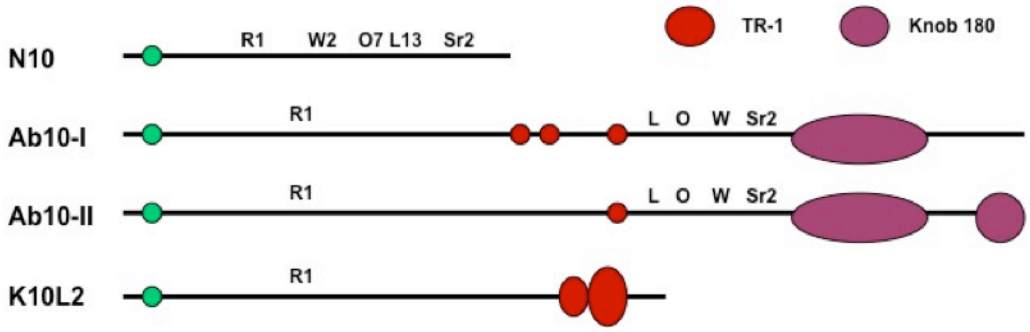

Figure 4.1: Chromosome 10 haplotype variants differ by knob content. Green circle represents the centromere and the long arm is enlarged and not to scale. R1 is a genetic marker for kernel color and marks the beginning of the abnormal haplotype. Ab10-I and K10L2 have TR-1 neocentromere movement and Ab10-II does not.

Table 4.1: Description of 32 samples used in analysis. “Rhoades” refers to lines isolated by Marcus Rhoades. Group A are genotypes with TR-1 neocentromere activity and Group B are genotypes without TR-1 neocentromere activity. Color alleles under Parental Cross include R (purple), r (colorless), R-ch (R-cherry), and R-st (R stippled).

Line	Geographic Origin	Subspecies	Haploype	Sample Number	Group	Parental Cross	Marker Genotyped
Rhoades	Outside Mexico City	mays	Ab10-I	3	A	RAb10I/RnjN10 X rN10/rN10	R allele, Kin618(all)
Rhoades	Outside Mexico City	mays	N10	3	B		
X233F	American Southwest	mays	Ab10-I	3	A	R-chAb10I/rN10 self	Kin618(all)
X233F	American Southwest	mays	N10	3	B		
K10L2	South Texas Inbred	mays	K10-L2	5	A	rK10L2/RnjN10 X RstN10/RstN10	r allele, C2RJJMR
K10L2	South Texas Inbred	mays	N10	3	B		
Rhoades	Unknown	mays	Ab10-II	3	B	RAb10II/RnjN10 X rN10/rN10	R allele, Kin618(all)
Rhoades	Unknown	mays	N10	3	B		
PI 566687	Mexico	parviglumis	Ab10-II	3	B	rN10/rN10 X rAb10II/rN10	Kin618(all)
PI 566687	Mexico	parviglumis	N10	3	B		

Table 4.2: Sequence files used in the transcriptome assembly and downstream analysis. R1 refers to forward read and R2 refers to the reverse read of the paired-end read set.

File Name	Total Reads	Average Read length	%GC	Plant Genotype	Group
A1-2.R1.fq.gz	14519485	35-76	49	Rhoades Ab10-I	A
A1-2.R2.fq.gz	14519485	35-76	51	Rhoades Ab10-I	A
A1-3.R1.fq.gz	15833903	35-76	50	Rhoades Ab10-I	A
A1-3.R2.fq.gz	15833903	35-76	51	Rhoades Ab10-I	A
A1-6.R2.fq.gz	17913093	35-76	50	Rhoades Ab10-I	A
A1-6.R2.fq.gz	17913093	35-76	51	Rhoades Ab10-I	A
A2-1.R1.fq.gz	12771889	35-76	50	Rhoades Ab10-II	B
A2-1.R2.fq.gz	12771889	35-76	51	Rhoades Ab10-II	B
A2-3.R1.fq.gz	12683268	35-76	50	Rhoades Ab10-II	B
A2-3.R2.fq.gz	12683268	35-76	51	Rhoades Ab10-II	B
A2-4.R1.fq.gz	15944619	35-76	50	Rhoades Ab10-II	B
A2-4.R2.fq.gz	15944619	35-76	51	Rhoades Ab10-II	B
AL2-1.R1.fq.gz	13836712	35-76	50	K10L2	A
AL2-1.R2.fq.gz	13836712	35-76	51	K10L2	A
AL2-3.R1.fq.gz	11810689	35-76	49	K10L2	A
AL2-3.R2.fq.gz	11810689	35-76	51	K10L2	A
AL2-4.R1.fq.gz	12144770	35-76	50	K10L2	A
AL2-4.R2.fq.gz	12144770	35-76	51	K10L2	A
AL2-5.R1.fq.gz	12968898	35-76	50	K10L2	A
AL2-5.R2.fq.gz	12968898	35-76	51	K10L2	A
AL2-6.R1.fq.gz	22274826	35-76	50	K10L2	A
AL2-6.R2.fq.gz	22274826	35-76	51	K10L2	A
K2-2.R1.fq.gz	9752492	35-76	50	*parviglumis* Ab10-II	B
K2-2.R2.fq.gz	9752492	35-76	51	*parviglumis* Ab10-II	B
K2-3.R1.fq.gz	10176440	35-76	49	*parviglumis* Ab10-II	B
K2-3.R2.fq.gz	10176440	35-76	50	*parviglumis* Ab10-II	B
K2-6.R1.fq.gz	12804250	35-76	50	*parviglumis* Ab10-II	B
K2-6.R2.fq.gz	12804250	35-76	51	*parviglumis* Ab10-II	B
N1-1.R1.fq.gz	7631444	35-76	49	Rhoades N10-I	B
N1-1.R2.fq.gz	7631444	35-76	50	Rhoades N10-I	B
N1-3.R1.fq.gz	14447476	35-76	49	Rhoades N10-I	B
N1-3.R2.fq.gz	14447476	35-76	51	Rhoades N10-I	B

N1-6.R1.fq.gz	19651677	35-76	50	Rhoades N10-I	B
N1-6.R2.fq.gz	19651677	35-76	51	Rhoades N10-I	B
N2-1.R1.fq.gz	14324115	35-76	50	Rhoades N10-II	B
N2-1.R2.fq.gz	14324115	35-76	51	Rhoades N10-II	B
N2-5.R1.fq.gz	11320166	35-76	49	Rhoades N10-II	B
N2-5.R2.fq.gz	11320166	35-76	50	Rhoades N10-II	B
N2-6.R1.fq.gz	10694418	35-76	49	Rhoades N10-II	B
N2-6.R2.fq.gz	10694418	35-76	51	Rhoades N10-II	B
NK2-4.R1.fq.gz	14196861	35-76	49	*parviglumis* N10	B
NK2-4.R2.fq.gz	14196861	35-76	50	*parviglumis* N10	B
NK2-8.R1.fq.gz	10843308	35-76	50	*parviglumis* N10	B
NK2-8.R2.fq.gz	10843308	35-76	51	*parviglumis* N10	B
NK2-9.R1.fq.gz	8526036	35-76	49	*parviglumis* N10	B
NK2-9.R2.fq.gz	8526036	35-76	50	*parviglumis* N10	B
NL2-2.R1.fq.gz	10074472	35-76	50	K10L2 N10	B
NL2-2.R2.fq.gz	10074472	35-76	51	K10L2 N10	B
NL2-5.R1.fq.gz	10461664	35-76	49	K10L2 N10	B
NL2-5.R2.fq.gz	10461664	35-76	50	K10L2 N10	B
NL2-3.R1.fq.gz	19566800	35-76	49	K10L2 N10	B
NL2-3.R2.fq.gz	19566800	35-76	51	K10L2 N10	B
X-A1.R1.fq.gz	11833305	35-76	50	X233F Ab10	A
X-A1.R2.fq.gz	11833305	35-76	51	X233F Ab10	A
X-A2.R1.fq.gz	19644615	35-76	50	X233F Ab10	A
X-A2.R2.fq.gz	19644615	35-76	51	X233F Ab10	A
X-A4.R1.fq.gz	13750408	35-76	49	X233F Ab10	A
X-A4.R2.fq.gz	13750408	35-76	51	X233F Ab10	A
X-N1.R1.fq.gz	11410162	35-76	49	X233F N10	B
X-N1.R2.fq.gz	11410162	35-76	52	X233F N10	B
X-N2.R1.fq.gz	12430476	35-76	50	X233F N10	B
X-N2.R2.fq.gz	12430476	35-76	51	X233F N10	B
X-N3.R1.fq.gz	12674662	35-76	49	X233F N10	B
X-N3.R2.fq.gz	12674662	35-76	50	X233F N10	B

Table 4.3: FPKM values across Group A for all contigs with FPKM values equal to 0 in Group B.

Gene ID	FPKM X_A1 X233F AB10-I Sample1	FPKM X_A2 X233F AB10-I Sample 2	FPKM X_A4 X233F AB10-I Sample 4	FPKM A1_2 Ab10-I Rhoades Sample2	FPKM A1_3 Ab10-I Rhoades Sample3	FPKM A1_6 Ab10-I Rhoades Sample 6	FPKM AL2_1 K10 L2 Sample1	FPKMA L2_3 K10L2 Sample3	FPKMA L2_4 K10L2 Sample4	FPKMA L2_5 K10L2 Sample5	FPKMA L2_6 K10L2 Sample6
c100009_g1	0	0	0	0	0	0	2.77	0	0	0	0
c100009_g4	0	0	1.67	0	0	0	0	0	0	0	0
c100027_g5	0	0	0	0	0	0	0	0	0	0	0
c100037_g2	0	0	0	0	0	0	1.7	0	1.03	0	4.55
c100061_g2	0	0	0	0	0	0	0	0	0	0	0
c100066_g2	0	0	0	1.94	0	0	0	0	0	0	0
c100006_g1	0.52	0	0	0	0	0.69	0.45	0.53	0	0	0
c100094_g2	0	0	0	0	0	0	0	0	0	0	0
c100097_g3	0	0	0	0	0	0	0	0	0	0	0
c100134_g2	0	0	0	0	0	0	1.7	0	0	0	0
c100197_g1	0	0	0	0	0	0	0	0.95	0	0	0
c100222_g7	0	0	0	0	0	0	0	0	1.82	0	0
c100243_g4	0	0	0	0	0	0	0	4.72	0	0	3.26
c100293_g2	0	0	0	0	0	0	0	0	0	0	0
c100331_g1	0	0	0	0	0	0	0.92	1.08	0	0	0
c100371_g3	0	0	0.86	0	0	0	0	0	0	0	0
c100371_g4	0	0	0	0	0	0	0	0	0	0	0
c100371_g5	0	0	0	0	0	0	0	0	0	0	0
c100389_g1	0	0	0	0	0	0	0	0	0	0	0
c100381_g1	0	1.04	0	0	0	0	0	0	0	0	0
c100441_g3	0	0	0	0	0	1.25	0	0	0	0	0
c100451_g1	0	0	0	0	0	0	0	0	0	0	0
c100539_g3	0	0	0	0	0	0	0	0	0	0	0
c100566_g1	0	0	0	0	0	0	0	0	0	0	0
c10056	0	0	0	0	0	0	0	0	0	0	0

8_g6											
c10059 1_g2	0	0	0	0	0	0	0	0	0	0	0
c10059 4_g2	0	0	0	0	0	0	0	0	0	0	3
c10064 0_g1	0	0	0	1.22	0.53	0.49	0	0.75	0	0	0
c10065 9_g2	0	0	0	0	0	0	0	0	0	0	0
c10072 0_g4	0	0	0	0	0	0	0	0	0	0	0
c10072 g1	0	0	0	0.91	0.79	1.44	0	0	0	0	0
c10078 9_g1	19.34	17.26	19.47	12.61	11.83	11.93	0	0	0	0	0
c10078 g1	0	0	0	0	0	1.18	1.58	0	0	0	0
c10082 g1	0	0.9	1.35	1.23	1.06	1.93	0	0	0	0	0.75
c10087 2_g2	0	0	0	0	0	0	0	0	0	0	0
c10088 1_g1	1.99	0	0	1.67	2.88	1.3	0	0	0	0	0
c10098 1_g2	0	0	0	0	0	0	0	0	0	2.14	0
c10099 2_g2	1.59	0	0	0	0	0	0	1.66	0	0	0
c10100 5_g4	0	0	0	0	0	0	0	0	0	0	0
c10101 2_g2	0	0	0	1.72	0	0	0	0	0	0	0
c10102 9_g4	0	0	0	0	0	0	0	1.84	0	0	0
c10115 4_g2	0.91	0	0.83	0.76	0	0	0	0	0	0	0.95
c10116 1_g5	1.97	0	0	0	0	0	0	0	0	0	0
c10116 g1	0	0	0	0	0	0	0	0	0	0	0
c10118 g1	0	0	1.39	1.27	0	0	0	0	1.63	0	0
c10126 1_g2	0	0.7	0.86	0.96	0	0	0	0	0	2.04	0
c10132 0_g4	0	0	0	0	0	0.64	0	0	2.06	0	0
c10135 9_g2	0	0.25	0.38	0	0	0	0	1.71	0.43	0.38	0.68
c10135 9_g3	0	0.69	0	0	0	0	0	0	0	0	0
c10138 3_g1	0	0	0	0	0	0	0	0	0	0	0
c10139 6_g1	0	0	2.12	0	0	0	0	2.38	0	0	0
c10145 0_g4	1.56	0	0	0	1.13	0	1.37	0	0	0	0
c10148 4_g1	0	1.36	4.12	0	0	1.44	0	0	0	0	0
c10148 g1	0	0	0	0	0	0	0	0	0	0	0
c1014_ g1	0	0	0	0	0	0	0	0	0	0	0
c10151 g1	0	0	0	0	0	0	0	0	0	0	0
c10153 g1	0	0	0	0	0	0	0	0	0	0	0
c10154 4_g2	0	0	0	0	0	0	0	0	0	0	0
c10154	0	0	0	0	0	0	1.1	2.56	0.86	2.63	2.25

7_g4											
c10157 7_g3	2.25	1.37	2.09	0	0	1.46	0	0	0	0	1.12
c10158 _g1	0	1.36	0	0	0	1.44	0	0	0	0	0
c10161 0_g2	0	0	0	2	1.72	1.55	0	0	0	0	0
c10162 7_g2	0	0.87	0	0	0	0	0	0	0	0	0
c10165 6_g3	0	0	0	0	0	0	0.72	0	0	0	0
c10166 3_g3	0	0	0.5	0	0	0	0	0	0	0	0
c10166 _g1	0	0.3	0	0	0	0	0	0	0	0	0
c10171 8_g4	0	0	0	0	0	0	0	0	0	0	0
c10175 4_g2	0	0	0	0	0	0	0	0	0	0	0
c10175 9_g4	0	0	0	0	0	0	0	0	0	0	0
c10176 3_g3	0	0	0.83	0	0	1.21	0	0	0	0	0
c10181 3_g3	0	0.7	0	0	0	0.75	1	0	0	0	0
c10181 3_g4	0.81	1.99	2.57	5.47	2.08	2.46	0	0	0	0	0
c10186 _g1	0	0	2.22	0	0	0	0	0	0	0	0
c10187 6_g3	0	0	0	0	0	0	0	0	0	0	0.46
c10198 8_g3	0	0	0	0	0	0	0	0	0	0	0
c10202 1_g1	0	0	1.78	0	0	0	0	0	0	0	0
c10203 6_g1	0	0.91	1.38	0	0	0	0	0	0	0	0
c10204 1_g2	0	0	0	0	0	0	0	0	0	0	0
c10204 _g1	8.15	1	0	0	0	0	0	0	0	0	0
c10208 _g1	0	0	0	0	0	0	0	0	0.28	0.25	4.46
c10217 8_g5	0	1.3	0	1.78	0	0	1.87	0	0	1.89	0
c10218 5_g4	0	0	0	0	0	0	0	0	0	0	0
c10221 1_g2	0	0	0	0	0	0	0	0	0	0	0
c10223 3_g2	0	0	0	0	0	0	0	0	0	0	0
c10227 _g1	1.78	1.09	0	0	0	0	0	0	0	0	0
c10229 3_g1	0	0	1.47	1.34	0	0	0	1.66	1.72	0	0
c10231 7_g3	3.79	0	0	0	0	0	0	0	0	0	0
c10232 5_g2	0	0	0	0	0	0	0	0	0	0	0
c10235 5_g5	0	0.79	0	0	2.82	1.71	0	0	0	0	0
c10237 g1	0	0	0	0	0	0	0.96	0	0	0.97	0
c10238 8_g3	0	0	0	0	0	0	0	0	0	0	0
c1023_ g1	0	1.12	0	0	0	0	0	0	0	0	0
c10242	0	0	0	9.45	8.2	10.71	0	0	0.4	0	0

2_g2											
c10242_g1	0	0	0	0	0	0	2.5	0	3.01	0	0
c10243_8_g3	0	0	0	0	0	0	0	0	0	0	0
c10247_0_g2	4.74	5.38	5.38	3.4	3.6	3.82	0	0	0	0	0
c10248_4_g3	0	0	0	0	0	0	0	0	0	0	0
c10251_4_g1	1.09	1.33	0	0	0	0.72	0	0	0	0	0
c10254_8_g2	0	0	0	2.35	0	0	0	0	0	0	0
c10256_4_g1	0	0	0	1.15	0	0	0	0	0	0	0.7
c10259_2_g1	0	0	0	0	0	0	0	3.48	0	0.44	1.82
c10259_7_g1	0	0	0	1.23	0.61	0.62	0	0	0	0	0
c10260_2_g2	0	0	0	0	0	0	0.89	0	0	0	0.53
c10260_g1	0	0	0	0	0	0	0	0	0	0	0
c10261_g1	0	0	0	0	0	0	0	0	0	0	0
c10265_7_g3	0	0	1.31	1.2	0	0	0	0	0	0	0
c10269_9_g5	0	0	0	0	0	0	0	0	0	0	0
c10270_8_g3	1.45	0.89	0	0	0	0	0	0	0	0	0
c10272_g1	1.23	0	0	0	0	0	0	0	1.32	1.1	0
c10274_9_g2	0	0	0	0	0	0	0	0	0	0	0
c10276_g1	0	0	2.39	0	0	0	0.77	0.9	0	0	0.92
c10277_5_g2	0	0	0	0	0	0	0	0	2.61	0	0.63
c10278_6_g4	1.09	0	0	0	0	0	0.96	0	0	0	0.57
c1027_g1	0	0.82	0	1.13	0	0	0	1.39	1.44	1.2	0
c10284_g1	0	0	0	0	0	0.54	0	0	0	0	0
c10286_0_g5	0	0	0	0	0	0	0	0	0	0	0
c10289_2_g2	0	0	0	0	0	0	0	0.38	0	0	0
c10289_3_g2	0	0	0	0	0	0	0	0	0	0	0
c10290_4_g2	0	1.66	0	0	0	0	0	1.41	0	0	0
c10291_4_g1	0	0	0	0	0	0	0	0	0	0	0
c10291_4_g3	0	0	0	0	0	0	0	0	0	0	0
c10291_5_g1	0	0	0	0	0	0	0	0	0	0	0
c10292_1_g1	0	0	0	0	0	0	0	0	0	2.05	0

Table 4.4: FPKM values across Group A for contigs that have FPKM > 0 for Ab10-I and FPKM = 0 for Ab10-II samples and N10 samples (all K10L2 samples left out of the analysis but shown in this figure).

Gene ID	**FPKM X_A1 X233F AB10-I Sample1**	**FPKM X_A2 X233F AB10-I Sample 2**	**FPKM X_A4 X233F AB10-I Sample 4**	**FPKM A1_2 Ab10-I Rhoades Sample2**	**FPKM A1_3 Ab10-I Rhoades Sample3**	**FPKMA1_6Ab10-I Rhoades Sample6**	**FPKM AL2_1 K10L2 Sample1**	**FPKMAL2_3 K10L2 Sample3**	**FPKMAL2_4 K10L2 Sample4**	**FPKMAL2_5 K10L2 Sample5**	**FPKMAL2_6 K10L2 Sample6**
c73250_g1	4.03	4.96	3.39	7.97	2.95	3.28	1.75	0	0	0	0
c77450_g2	2.94	3.6	2.84	3.54	3.69	2.85	0.19	0	0	0	0

Table 4.5: Contigs from the transcriptome assembly with over 9.9 log fold change difference between Group A and Group B. logFC is the log fold change of FPKM values between Group A and Group B, logCPM is the log counts per million, or reads mapping to the gene in each group normalized by per-group library coverage, P value represents the significance of the differential expression, based on both logFC and logCPM, and FDR is the false discovery rate. BLASTX result is the contig nucleotide sequence aligned using BLASTX to the NCBI non-redundant protein database and e value is the significance of that BLASTX result.

Gene ID	logFC	logCPM	PValue	FDR	BLASTX Result	BLASTX e value
c99545_g1	12.30876913	5.218679728	5.10E-17	1.16E-12	PIMT_WHEAT Protein-L-isoaspartate O-methyltransferase	1.00E-21
c103880_g1	12.11909137	5.03021417	1.84E-16	2.40E-12	MYCA_BACIU Mycosubtilin synthase subunit A	3.1
c85882_g2	11.52254979	4.438703797	1.00E-14	4.98E-11	FRD3_ARATH MATE efflux family protein FRD3	4.00E-14
c90689_g1	11.49877671	4.415177237	1.17E-14	4.98E-11	RPM1_ARATH Disease resistance protein RPM1	4.00E-58
c89140_g1	11.49877671	4.415177237	1.17E-14	4.98E-11	HARB1_HUMAN Putative nuclease HARBI1	4.00E-04
c85148_g1	11.42921107	4.346356563	1.86E-14	7.03E-11	PK1IP_CHICK p21-activated protein kinase-interacting protein 1-like	9.9
c102112_g1	11.41234341	4.329674964	2.08E-14	7.45E-11	COB21_ORYSJ Coatomer subunit beta'-1	3.00E-51
c88116_g1	11.33389894	4.252124729	3.50E-14	1.13E-10	ADHL6_ARATH Alcohol dehydrogenase-like 6	0.008
c96436_g1	11.09583952	4.017093779	1.74E-13	4.69E-10	MERD_SALTI HTH-type transcriptional regulator MerD	5.8
c104756_g1	11.0799053	4.001380487	1.93E-13	4.87E-10	TF29_SCHPO Transposon Tf2-9 polyprotein	2.00E-92
c95447_g1	11.00312615	3.925700393	3.20E-13	6.80E-10	YI31B_YEAST Transposon Ty3-I Gag-Pol polyprotein	2.00E-113
c100254_g2	10.9746837	3.897680002	3.85E-13	7.95E-10	C7A15_ARATH	4.00E-177

					Cytochrome P450 72A15	
c75048_g1	10.91605923	3.839951839	5.66E-13	1.10E-09	FYV10_EMENI Protein fyv10	1.7
c84146_g1	10.9100636	3.834049922	5.89E-13	1.11E-09	COPIA_DROME Copia protein	0.24
c95182_g1	10.87355695	3.798122256	7.47E-13	1.37E-09	SUS1_ARATH Transcription and mRNA export factor SUS1 {ECO:0000255|HAMAP-Rule:MF_03046}	1.1
c104011_g1	10.77143299	3.697696967	1.45E-12	2.36E-09	Unknown	n/a
c89750_g1	10.75145303	3.678063453	1.66E-12	2.62E-09	5NT3B_HUMAN 7-methylguanosine phosphate-specific 5'-nucleotidase	1.00E-51
c94391_g1	10.58079417	3.510564714	5.00E-12	7.24E-09	PALC_EMENI pH-response regulator protein palC	4.8
c92950_g2	10.53477735	3.46546495	6.72E-12	9.00E-09	VGFR2_MOUSE Vascular endothelial growth factor receptor 2	6.4
c100490_g2	10.495276	3.426774101	8.66E-12	1.09E-08	OPT5_ARATH Oligopeptide transporter 5	7.00E-10
c80062_g1	10.47916756	3.411002501	9.61E-12	1.15E-08	SYV_PSEHT Valine--tRNA ligase {ECO:0000255|HAMAP-Rule:MF_02004}	8.8
c60672_g1	10.42132802	3.354403352	1.39E-11	1.55E-08	ANM61_ORYSI Probable protein arginine N-methyltransferase 6.1	0.012
c73894_g1	10.32547593	3.260717531	2.56E-11	2.61E-08	unknown	n/a
c93391_g3	10.30734281	3.24301036	2.88E-11	2.74E-08	FKB65_ARATH Peptidyl-prolyl cis-trans isomerase FKBP65	0.35
c67487_g1	10.30734281	3.24301036	2.88E-11	2.74E-08	POLX_TOBAC Retrovirus-related Pol polyprotein from transposon TNT 1-94	0
c99891_g3	10.29819007	3.234074611	3.05E-11	2.81E-08	TPM1_LIZAU Tropomyosin alpha-1 chain	0.43
c95576_g1	10.18357323	3.122292605	6.30E-11	5.23E-08	GAUT1_ARATH Polygalacturonate 4-alpha-galacturonosyltransferase	0.03
c84983_g2	10.15343747	3.092939552	7.62E-11	5.83E-08	LOX6_ORYSJ Probable lipoxygenase 6	5.00E-07
c103906_g2	10.15343747	3.092939552	7.62E-11	5.83E-08	POL3_DROME Retrovirus-related Pol polyprotein from transposon 17.6	1.00E-75
c100485_g2	10.14325069	3.083020987	8.12E-11	6.07E-08	PHYC_SORBI Phytochrome C	4.00E-79
c104100_g2	10.12265876	3.06297694	9.25E-11	6.63E-08	NALCN_RAT Sodium	5.9

					leak channel non-selective protein	
c101131_g2	10.12265876	3.06297694	9.25E-11	6.63E-08	Unknown	n/a
c88030_g1	10.09120907	3.032378918	1.13E-10	7.91E-08	IGS10_RAT Immunoglobulin superfamily member 10	7.3
c104773_g1	10.0590585	3.001117952	1.38E-10	9.57E-08	POL3_DROME Retrovirus-related Pol polyprotein from transposon 17.6	4.00E-39
c103263_g1	10.03721971	2.979894587	1.58E-10	1.08E-07	unknown	n/a
c96687_g1	10.01504525	2.958354387	1.81E-10	1.22E-07	MAOX_PHAVU NADP-dependent malic enzyme	4.00E-22
c77870_g1	9.981131119	2.925428897	2.24E-10	1.41E-07	CRYAB_ANAPL Alpha-crystallin B chain	0.48
c101919_g2	9.958070529	2.903053623	2.59E-10	1.59E-07	SPL1_ORYSJ Squamosa promoter-binding-like protein 1	2.00E-24
c86837_g1	9.958070529	2.903053623	2.59E-10	1.59E-07	GSO2_ARATH LRR receptor-like serine/threonine-protein kinase GSO2	5.00E-04
c86505_g1	9.922773416	2.868826376	3.22E-10	1.87E-07	COPIA_DROME Copia protein	3.00E-16

Table 4.6: Differential expression data for all annotated kinesins in the *de novo* transcriptome assembly with significant differential expression between Group A and Group B (P value < 0.05). BLASTX result is the contig nucleotide sequence aligned using BLASTX to the NCBI non-redundant protein database and e value is the significance of that BLASTX result. logFC is the log fold change of FPKM values between Group A and Group B, logCPM is the log counts per million of reads mapping back across all 32 groups, P value represents the significance of the differential expression, based on both logFC and logCPM, and FDR is the false discovery rate.

Gene ID	BLASTX Result	e value	logFC	logCPM	PValue	FDR
c95145_g1	ATK5_ARATH Kinesin-5	1.00E-04	5.265093611	1.487384235	1.34E-05	0.001238473
c80658_g1	KP1_ARATH Kinesin KP1	2.00E-13	-7.64864474	0.75836958	7.45E-05	0.005124384
c82958_g1	NACK1_TOBAC Kinesin-like protein NACK1	3.00E-117	-3.240863773	1.454881111	0.000851461	0.036040401
c95588_g1	K125_TOBAC 125 kDa kinesin-related protein	0.12	-3.215595134	1.070362732	0.002088451	0.071358567
c98366_g1	POK2_ARATH Phragmoplast orienting kinesin 2	2.4	-2.137843379	1.900381929	0.010249229	0.226417818
c78568_g1	KINH_NEUCR Kinesin heavy chain	2.00E-19	-1.809478298	1.655432733	0.032386763	0.48294751
c123802_g1	KLC_STRPU Kinesin light chain	0.3	5.835270599	-0.63897125	0.034782609	0.48294751
c90974_g1	CTK2_XENLA Carboxy-terminal kinesin 2	4.00E-25	-1.602341119	2.009058112	0.045254176	0.582696746
c90277_g1	KI26L_CAEBR Kinesin-like protein vab-8	7.3	1.703998728	1.574356794	0.046667979	0.595031561

Table 4.7: Contigs generated by the de novo transcriptome analysis that align using BLASTn to Kin618 (B5) with an e value under 0.001. BLASTN B73 location, percent identity, and e value all represent the contig subsequently compared to the B73 genome using BLASTn on maizegdb.org. Annotated gene at B73 Blastn location represents the B73 annotation of that particular genomic location. The differential expression values for each contig is next represented: logFC is the log fold change of FPKM values between Group A and Group B, logCPM is the log counts per million of reads mapping back across all 32 groups, P value represents the significance of the differential expression, based on both logFC and logCPM, and FDR is the false discovery rate.

Gene ID	BLASTN B73 location	Percent Identity	e Value	Annotated Gene at B73 BLASTn Location	logFC	logCPM	PValue	FDR
Kin618					-0.00430405	-1.513773068	1	1
c128879_g1	chr7	80.61	8.460e-23	Kinesin motor domain	n/a	n/a	n/a	n/a
c103505_g1	chr8	99.27	6.329e-137	Magnesium chelatase	0.632290671	4.051463171	0.36353501	1
c87559_g1	chr7	97.42	0	kinesin	0.381257688	1.545283922	0.688138494	1
c98269_g1	chr1	100	0	DEAD/DEAH box helicase domain containing protein	0.460280326	6.608133495	0.486070295	1
c63227_g1	chr10	98.98	0	no gene	-1.469454998	-1.098836281	0.658385093	1
c41197_g1	chr10	98.84	0	no gene	-0.004557229	-1.098735407	1	1
c104556_g1	multiple locations				-1.299380755	0.544786546	0.219276674	1
c98269_g3	chr7	99.73	0	kinesin	-0.252424242	-0.191888356	1	1
c11781_g	chr10	98.8	0	no gene	n/a	n/a	n/a	n/a

1								
c120764_g1	chr10	84.08	3.914e-56	no gene	n/a	n/a	n/a	n/a
c92307_g1	chr2	94.67	0	kinesin	0.502960257	6.134088333	0.448367745	1
c104294_g1	chr2	99.7	1.637e-170	kinesin motor domain containing protein, kinesin-like calmodulin-binding protein	-0.253720406	6.131079852	0.70299739	1
c104259_g1	chr2	99.55	0	Tetratricopeptide repeat (TPR)-like superfamily protein	-0.01763051	4.504877315	0.982153142	1
c104508_g1	chr3	98.37	0	Kinesin motor domain	-0.658365738	2.682239445	0.374611498	1
c58191_g1	chr3	85.47	1.406e-149	non-annotated gene	-7.465200232	0.600898365	0.000164799	0.009603348
c48401_g1	chr3	99.47	8.882e-93	kinesin motor domain	-4.085186251	-1.513878327	0.523809524	1
c103027_g1	chr1	100	0	Kinesin motor domain	-0.55960967	5.696862035	0.402835104	1
c101104_g1	chr6	99.72	0	Kinesin motor domain	0.2156862	2.327646939	0.809097965	1

CHAPTER 5: CONCLUSIONS AND DISCUSSION

Meiosis is a complicated process with a straightforward outcome: equal segregation of genes. Abnormal chromosome 10 (Ab10-I) in maize segregates among offspring at levels between 70-80%, far above the 50% predicted by Mendel [1]. To accomplish this feat, Ab10-I exploits complicated meiotic machinery and asymmetric female meiosis in a process called Meiotic Drive. Ab10-I is a haplotype variant of normal chromosome 10 (N10) distinguished by the presence of clusters of dense tandem repeats called knobs. Knobs are composed of a 180-bp repeat sequence (Knob 180), a 350-bp repeat sequence (TR-1), or sometimes both [2,3]. When Ab10-I is present, knobs form neocentromeres and race laterally along the meiotic spindle to the poles of the dividing cell. In maize female meiosis, only the bottom, or most polar, cell of the four products develops into the seed. Neocentromeres pull the abnormal chromosome into this cell and cause meiotic drive [4].

Abnormal chromosome 10 has been studied since its discovery in 1942 yet prior to the work presented here, nothing was known about the molecular mechanism of drive. Previous studies determined that knobs were activated by trans-acting loci located in sequence unique to Ab10 [5]. Subsequent work determined a clear association between Knob 180 neocentromeres and meiotic drive, whereas TR-1 neocentromere activity is not sufficient to cause drive [6,7]. This dissertation offers the first unique meiotic transcriptome and genomic Ab10 sequence as well as a list of meiotically expressed genes unique to the Abnormal haplotype that likely play a role in drive. Of the nine gene candidates, one is a C-terminal kinesin belonging to a multi-copy gene

family present on the "distal tip" of Ab10 we call Kin618. Our mutational and cytogenetic analysis of Kin618 concludes that it is responsible for moving Knob 180 neocentromeres and causing preferential segregation of the selfish chromosome, offering the first molecular evidence of Ab10 drive.

A preliminary genetic map of the abnormal chromosome was previously created using a series of deficiency lines, or versions of Ab10 with breakpoints along the long arm of the chromosome [8]. The smallest deficiency, Df(L), contains all of the abnormal chromosome except for the distal euchromatic tip with the loci causing preferential segregation and Knob 180 neocentromere activity [7,9]. By using a comparison transcriptome approach we successfully identified nine genes on the distal tip of Ab10 expressed during meiosis. Six of the nine genes are conserved across geographic boundaries and three are meiotic-specific. Genomic sequence from a Bacterial Artificial Chromosome (BAC) library that corresponds to the nine distal genes allowed us to construct the first genetic map of the distal tip of Ab10. The distal tip contains at least one duplication and is composed of an amalgamation of other parts of the maize genome. Though past studies have hypothesized that Ab10 is "alien" DNA and the result of a wide species cross, we instead show that Ab10 arose during a time of genome reshuffling, perhaps after the last whole genome duplication within the maize lineage [10].

A suite of Ab10-I mutants combined with the nine candidate genes identified in Chapter 2 allowed us to finally test for a molecular mechanism of drive. The five suppressor of meiotic drive (*smd*) mutants no longer preferentially segregate (*smd1*, *smd3*, *smd8*, *smd12*, and *smd13*) and were created using Robertson's Mutator (*Mu*), a DNA transposon [9,11]. Though extensive effort was spent in search of novel *Mu* insertions in these five lines, it came to naught and the *smd* mutants remained uncharacterized. By employing PCR markers developed from the nine

distal genes, *smd3* and *smd8* were finally revealed as large terminal deletions, with *smd3* lacking seven of the distal tip genes. Meiotic expression analysis revealed that both gene 430 and gene 248 are downregulated in mutants *smd1*, *smd12*, and *smd13,* providing potent evidence that these two genes are involved in the drive mechanism.

Interestingly, of the nine distal tip genes only one codes for a fully function protein: a C-terminal kinesin we call Kin618. Kin618 belongs to the Kinesin-14A family containing members including Drosophila Ncd, mouse KIFC1, and Arabidopsis ATK1 as well as maize Kinesin 11 (Kin11) [12,13]. Kin618 and Kin11 share 88% homology at the nucleotide level along their coding sequence and differ only in their first exon in the "tail" region, at which a kinesin will contact its cargo. Unlike Kin11, Kin618 is part of a multi-copy gene family with approximately six members, all located on the distal tip of Ab10. Chapter 3 characterizes three of these: C4, B5, and B2, the last of which is likely not transcribed. B5 and C4 differ in several SNPs located in the kinesin tail region and B5 features an alternatively spliced first exon. This evidence suggests that Kin618 either functions as a heterodimer or there are multiple, functionally distinct holoenzymes.

Kin618 is the perfect candidate to grab onto Knob 180 neocentromeres and pull them along microtubules into the predestined cell. The most convincing evidence that Kin618 causes drive comes from one of the five Ab10-I mutants, *smd12*. The *smd12* line is an epimutant, in which Kin618 expression has been silenced by an increase in DNA methylation. Like all the other *smd* mutants, during meiosis *smd12* does not form Knob 180 neocentromeres but still forms TR-1 neocentromeres. Targeted bisulfite sequencing revealed increases in CG, CHG, and CHH methylation in the 5'UTR of both the C4 and B5 copies of Kin618. The fact that all Kin618 gene copies have been silenced in *smd12* points to a homologous-sequence based

methylation mechanism, such as RdDM. It is unclear why *Mu* mutagenesis led to an epimutant. A native *Mu* element directly upstream of Kin618 B5 prompts the theory that the methylation machinery was silencing this native *Mu* that had been activated by the mutagenesis screen. The methylation marks may have subsequently spread downstream into the promoter of Kin618 B5 and silenced the rest of the Kin618 copies due to homology. Conversely, the epimutant may have nothing to do with *Mu*, but rather have been induced by a small structural change, such as an inversion, near a copy of Kin618 that then prompted the RdDM machinery. *Mu* mutagenesis is not a gentle process as both *smd3* and *smd8* are large terminal deletions. Nevertheless, *smd12* is a potent example of an experimentally induced plant epimutation that silences multiple gene copies. Further characterization of this mutant line will reveal much about the relationship between epigenetics, transposons, and gene expression.

This thesis presents the discovery and characterization of the Knob 180 neocentromere motor, Kin618, yet the purpose and movement mechanism of TR-1 knobs remain a puzzle. Though Chapter 2 and Chapter 3 focus on Ab10-I, there are four characterized haplotype variants of the maize 10th chromosome that vary in knob content, drive ability, and neocentromere type activation: Ab10-I, Ab10-II, Ab10-III, and K10L2 [6,14]. Ab10-II and Ab10-III preferentially segregate and have both TR-1 and Knob 180 repeats. Notably, Ab10-II does not activate TR-1 neocentromeres. K10L2 does not have true drive but instead segregates at a statistically significant 51%. K10L2 only has TR-1 repeats and lacks both Kin618 and Knob 180 neocentromere activity. When paired with either Ab10-I or Ab10-II in a single plant, K10L2 suppresses the drive of these two abnormal haplotypes. This suppressive phenotype led previous studies to conclude that TR-1 and Knob 180 are locked in an arms-race like struggle to proliferate across the maize genome [6].

This theory of competition by suppression led us to hypothesize that a second kinesin specific to the TR-1 repeats exists on both Ab10-I and K10L2. As Ab10-II lacks TR-1 neocentromere activity, it also lacks this putative TR-1 gene. To test this hypothesis, we conducted a second meiotic transcriptome comparison experiment between genotypes that activate TR-1 neocentromeres (Ab10-I and K10L2) and those that do not (Ab10-II and N10). By constructing a de novo assembly of the haplotype variant transcriptome and calculating expression values for each contig across 32 plants and 10 genotypes, we searched for this elusive kinesin. A differential expression analysis between Group A, containing TR-1 neocentromere movement, and Group B, lacking TR-1 neocentromere movement, presented no convincing candidates shared across all the plants we incorporated in our study. Annotation of the transcriptome allowed us to focus on every single annotated kinesin contig, numbering at 134, and their expression values across all 32 samples. Again, none of the annotated kinesins showed any expression variation between Group A and Group B.

Though Kin618 clearly stands out in any abnormal transcriptome analysis, there is no corresponding candidate for the TR-1 motor. The top 40 differentially expressed candidates between genotypes with TR-1 neocentromere activity and those without are not specific to all members of one group and are mostly transposable elements that are highly expressed in a few members. It may be that TR-1 neocentromeres employ a native maize kinesin that gains specificity by non-coding RNA, which would not be detected by in our analysis. Past phylogenetic comparisons among all the Ab10 variants based on loci mapping near the TR-1 gene conclude that K10L2 is more closely related to N10 than Ab10-I or Ab10-II [14]. It could also be that whatever gene activates TR-1 is so closely related to the N10 homologue that our de novo transcriptome assembly interpreted them as the same transcript.

The meiotic drive of maize Abnormal chromosome 10 offers an unparalleled chance to understand true meiotic drive and its effect on an eukaryotic genome. By identifying nine genes specific to the abnormal haplotype, as well as a highly specific abnormal kinesin, the research presented here is the first to shed light on the molecular mechanism of the selfish chromosome. Furthermore, the *smd12* epimutation joins a relatively rare group of induced plant epimutants. The meiotic transcriptomes of 32 geographically diverse maize lines generated in Chapter 4 presents highly comprehensive expression data of chromosome 10 variants and will be useful in future analyses. Most importantly, the identification and characterization of the multi-copy gene family of kinesin Kin618 offers exciting new evidence of how a gene develops a novel cellular function and hijacks canonical chromosome movement for its own selfish propagation.

References

1. Rhoades, M. M. Preferential Segregation in Maize. *Genetics* **27**, 395–407 (1942).
2. Ananiev, E. V, Phillips, R. L. & Rines, H. W. A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10785–90 (1998).
3. Dennis, E. S. & Peacock, W. J. Knob heterochromatin homology in maize and its relatives. *Journal of molecular evolution* **20**, 341–50 (1984).
4. Rhoades, M. Preferential segregation in maize. *Heterosis* 66–80 (1952).
5. Hiatt, E. N., Kentner, E. K. & Dawe, R. K. Independently regulated neocentromere activity of two classes of tandem repeat arrays. *The Plant cell* **14**, 407–20 (2002).
6. Kanizay, L. B., Albert, P. S., Birchler, J. A. & Dawe, R. K. Intragenomic conflict between the two major knob repeats of maize. *Genetics* **194**, 81–9 (2013).
7. Kanizay, L. B. The variants of maize chromosome 10 and their roles in meiotic drive. 111 (2011).
8. Mroczek, R. J., Melo, J. R., Luce, A. C., Hiatt, E. N. & Dawe, R. K. The maize Ab10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. *Genetics* **174**, 145–54 (2006).
9. Hiatt, E. N. & Dawe, R. K. Four loci on abnormal chromosome 10 contribute to meiotic drive in maize. *Genetics* **164**, 699–709 (2003).
10. Wei, F. *et al.* Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLoS genetics* **3**, e123 (2007).
11. Dawe, R. K. & Cande, W. Z. Induction of centromeric activity in maize by suppressor of meiotic drive 1. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8512–7 (1996).
12. Lawrence, C. J. *et al.* A standardized kinesin nomenclature. *The Journal of cell biology* **167**, 19–22 (2004).
13. Miki, H., Okada, Y. & Hirokawa, N. Analysis of the kinesin superfamily: insights into structure and function. *Trends in cell biology* **15**, 467–76 (2005).
14. Kanizay, L. B. *et al.* Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in Zea mays. *Heredity* **110**, 570–7 (2013).