Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Gelsey Jaymes April 10, 2023

Evolution of Centromeric Protein Coding Genes in the *Mus* Genus

by

Gelsey Jaymes

Dr. Jitendra Thakur
Adviser

Biology

Dr. Jitendra Thakur

Adviser

Dr. Michal Arbilly

Committee Member

Dr. Arri Eisen

Committee Member

Dr. Mark Risjord

Committee Member

2023

Evolution of Centromeric Protein Coding Genes in the *Mus* Genus

By

Gelsey Jaymes

Dr. Jitendra Thakur

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2023

Abstract

Evolution of Centromeric Protein Coding Genes in the *Mus* Genus

By Gelsey Jaymes

Centromeres are chromosomal loci that facilitate error-free chromosome segregation during mitosis and meiosis to avoid aneuploidy and genetic instability, hallmarks of cancers and chromosomal disorders. Paradoxically, while the essential centromere function is strongly conserved, its components (centromeric DNA and associated proteins) are rapidly evolving. Centromeric components provide a unique system to understand evolutionary selection pressure acting on essential genomic components. Our lab has found that genes encoding centromeric proteins are under adaptive Darwinian selection in several *Mus* species. My study involves investigating the patterns of evolutionary changes in centromeric protein-coding genes in *Mus* species and how they contribute to the overall evolutionary selection pressure observed in the *Mus* genus. I found that centromeric proteins evolve differently in different phylogenetic branches of the *Mus* genus. Furthermore, centromeric proteins show different amounts of localization to centromeres in *Mus* species suggesting functional evolution in progress. Additionally, toward identifying possible copy number variations, which might be responsible for the rapid evolution of these genes, I have developed a cytological approach to determine the copy number of one of the centromeric protein-coding genes. Understanding the evolution of centromeric protein-coding genes will provide greater insight into centromere function and chromosomal disorders.

Evolution of Centromeric Protein Coding Genes in the *Mus* Genus

By

Gelsey Jaymes

Dr. Jitendra Thakur

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2023

Acknowledgements

I would like to recognize Dr. Thakur for her continual direction and support alongside a wonderful laboratory learning opportunity. I would like to acknowledge Olivia Morrison for further mentorship, guidance, and collaboration, Jenika Packiaraj and Nachiket Patil for technical troubleshooting and the rest of the Thakur laboratory for their help and encouragement. Lastly, I would like to acknowledge the committee members for their time and efforts in support of my thesis.

Table of Contents

Introduction

Centromeres determine where spindle microtubules attach on a given chromosome and are vital for chromosome segregation during cell division (Bloom and Costanzo 2017). Centromeres, with associated centromeric protein complexes, are essential for error-free chromosome segregation during mitosis and meiosis to avoid chromosome division errors such as aneuploidy and genetic instability, hallmarks of cancers and chromosomal disorders. Within eukaryotic chromosomes, DNA is packaged by histone proteins into chromatin. The basic unit of chromatin is the nucleosome, which consists of 146 bp long DNA wrapped around an octamer of histones (2 copies of each histone H2A, H2B, H3 and H4). At centromeres, nucleosomes are slightly different – histone H3 is replaced by a variant of H3 called CENP-A. CENP-A is the marker of centromere identity and function in most eukaryotes. CENP-A is deposited to centromeres by a specific chaperone called HJURP, which associates with chromatin during the late telophase/early G1 phase. Genes with essential functions tend to remain conserved because mutations disrupt their dedicated essential function. Therefore, given their essential function in chromosome segregation, it was previously assumed that centromeres and their associated components would be evolutionarily conserved. However, both centromeric DNA loci and several centromeric proteins were found to be rapidly evolving in eukaryotes. (Sanchez-Pulido et al. 2009, Kumon et al. 2021).

The centromere drive model, derived from the meiotic drive model, has been proposed to explain the rapid evolution of centromeric components. Female meiosis is asymmetric; out of four meiotic products, only one gives rise to the egg, while the other three are discarded as polar bodies. The meiotic drive model suggests that a variant gene that somehow can get access to the

egg is preferentially transmitted to future generations. Such an advantage in preferential transmission will allow the variant to spread across the population rapidly. For example, an increase in the size of the centromeric DNA locus can lead to increased microtubule binding, which could provide advantages to this centromere variant in segregating to the egg during female meiosis. However, such a change will pose fitness cost in males where meiosis is symmetrical. According to the centromere drive model, centromeric proteins evolve to suppress the harmful effects of centromeric DNA evolution in males. It remains to be proven whether centromeric proteins indeed evolve to functionally counter the evolution of centromeric DNA. Therefore, understanding evolutionary selection pressures and the functional consequences of the rapid evolution of centromeric protein-coding genes is of paramount interest to both evolutionary biology and functional genomics. Moreover, centromeric proteins, including HJURP, are overexpressed in several types of cancer (Dunleavy et al. 2009). A better understanding of HJURP protein evolution may also provide insight into the role of centromere mutations in cancer (Bloom and Costanzo 2017), especially alongside other centromeric proteins of the centromere complex.

My aim is to understand the how the evolution of essential centromeric proteins, including CENP-A, CENP-B, CENP-C, and CENP-T, affect centromere function in evolving species (Figure 1).

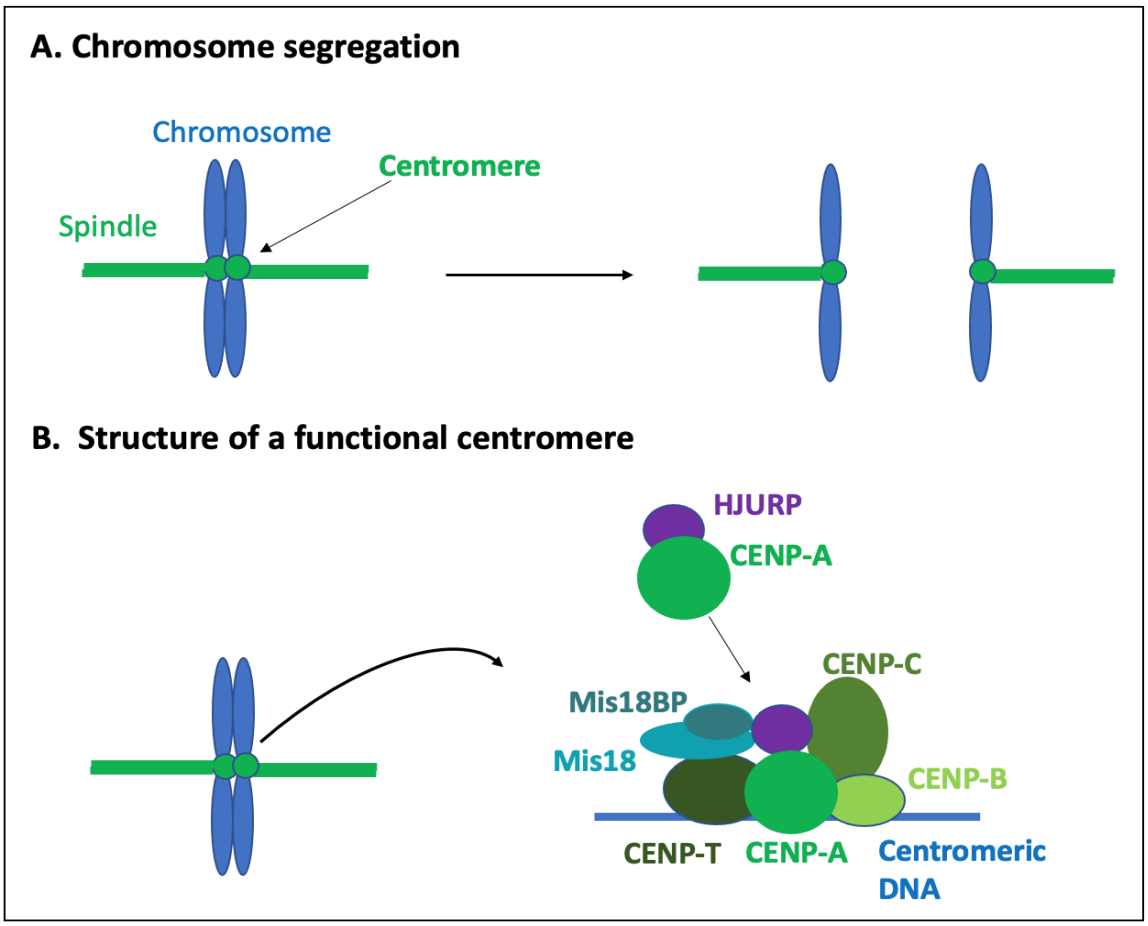


Figure 1. Functional centromeres facilitate chromosome segregation during cell division A. Spindle microtubules capture chromosomes by binding to functional centromeres and pull sister chromatid toward opposite poles to allow for chromosomes segregation. B. Components of a functional centromere. Centromeric proteins are localized to and bind centromeric DNA. HJURP localizes CENP-A to the centromeric DNA region, which allows for binding of additional proteins, including CENP-B, CENP-C, CENP-T, Mis18 and Mis18BP that are necessary for the attachment of the centromere to spindles. Together, centromeric proteins and the centromeric DNA form functional chromosome segregation machinery (Perpelescu, 2011).

Results

1. Centromeric protein-coding genes are evolving differently across different phylogenetic branches of the *Mus* genus.

The *Mus* genus is one of the best studied mammalian groups, and whole genome sequencing data is readily available for several closely related species. Various *Mus* lineages and species allow

for efficiently tracking centromeric protein evolutionary changes between species with varying but known evolutionary distances and relatedness (Figure 2).

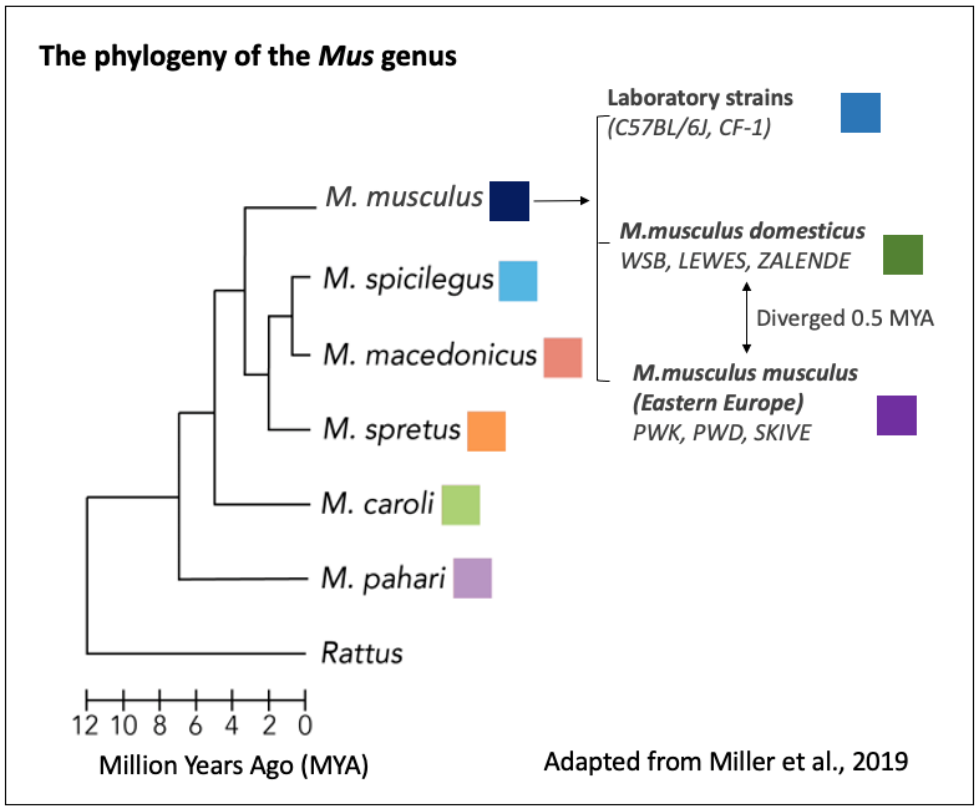


Figure 2. The phylogeny of the *Mus* genus. The *Mus* phylogenetic tree showing early diverging (separated by < 0.5 MYA) and late diverging lineages/species for which publicly available Whole Genome or transcriptome datasets are available for evolutionary analysis of protein-coding genes. Adapted from Miller et al. 2020.

Previous work from our lab found that HJURP and CENP-C are under long-term and rapid, short-term adaptive evolution in several *Mus* species (Morrison et al., unpublished). Adaptive evolution allows for the propagation of evolutionary changes that are beneficial to evolving species or lineages. Such adaptive and rapid evolution makes centromeric proteins excellent candidates for investigating evolutionary forces and their functional consequences on essential genes. However, two species, *Mus spicilegus* and *M. macedonicus* from one of the key *Mus* phylogenetic branches,

were not included in this study due to their absence from the assembled genome database at Ensembl.

My goal was to determine how centromeric protein-coding genes are evolving in the *M. spicilegus* and *M. macedonicus* phylogenetic branch with respect to the remaining *Mus* genus. Toward this, I extracted centromeric protein-coding genes and transcript sequences from publicly available whole genome and transcriptome datasets (Kodama, 2012 and SRA/Miller et al. 2020). I first extracted *M. spicilegus* transcripts for centromeric proteins based on homology to its close relative the Algerian mouse, *M. spretus* (phylogenetic distance shown in Figure 2) (Couger et al. 2018). Subsequently, by utilizing the constructed *M. spicilegus* transcripts, I extracted the centromeric protein transcripts for its close relative *M. macedonicus*. Finally, I compared centromeric protein-coding transcripts and found that evolutionary changes were concentrated in specific regions across centromeric protein-coding genes in *M. spicilegus* and *M. macedonicus* (Figure 3A).

Next, I investigated the short-term early selection pressure by comparing the mutation rates between pairs of species. This was performed by calculating the dN/dS ratio that compares the rate of nonsynonymous amino acid changes (dN) to the rate of synonymous amino acid changes (dS) between closely related species (Booker et al. 2017). In neutrality, where selection is not occurring, dN/dS should be equal to 1. The dN/dS ratio greater than 1 indicates positive selection where a beneficial nonsynonymous mutation sweeps through the population and becomes fixed. This means that the frequency of this allele increases in the population. The dN/dS ratio less than 1 indicates negative selection in which the harmful amino acid changes are discarded i.e., purifying selection. I determined if centromeric protein-coding genes are undergoing positive or

negative selection in *M. spicilegus* and *M. macedonicus,* and other closely related *Mus* species (Figure 3B). I found that among centromeric proteins analyzed, only CENP-T was undergoing positive selection between *M. m. domesticus* and *M. macedonicus*. Next, I included *M. spicilegus* and *M. macedonicus* in the long-term adaptive selection analysis previously performed in the lab using CodeML software. The CodeML analysis demonstrates long-term evolutionary trends through a comparison of many selected species, with varying evolutionary distances and relatedness, in the *Mus* genus. As previously reported in the lab, I found that HJURP and CENP-C were undergoing positive selection (Figure 3C). These results suggest that the early evolution of CENP-T in the *M. macedonicus and M.m.domesticus* species pair is not adaptive across the long evolutionary distance in the *Mus* genus. Furthermore, the long-term adaptive evolution of HJURP and CENP-C observed in the *Mus* genus is absent in the *M. spicilegus* and *M. macedonicus* branch. Overall, these findings suggest that centromeric protein-coding genes are evolving differently across different branches of the *Mus* genus.

A. Clustering of evolutionary changes across centromeric protein-coding genes

B. Short-term selection between *Mus* species pairs

	macedonicus vs Mmd	*macedonicus vs Mmm*	*macedonicus vs pahari*	*macedonicus vs caroli*	*macedonicus vs spretus*	*macedonicus vs spicilegus*
HJURP	0.466	0.765	0.602	0.484	0.596	0.577
CENPA	n/a (dN=0)	n/a (dN=0)	0.299	0.141	n/a (dN=0)	n/a (dN=0)
CENPB1	0.207	0.177	0.117	0.207	0.177	0.311
CENPB2	0.310	0.271	0.143	0.310	0.271	0.310
CENPC	0.635	0.516	0.516	0.782	0.544	0.454
CENPT	1.051	0.843	0.452	0.410	0.777	0.341
Mis12	0.098	0.250	0.155	0.030	0.125	0.248
Mis18a	0.149	0.148	0.399	0.315	0.098	0.301
Mis18bp1	0.687	0.677	0.579	0.749	0.665	0.690

C. Long-term selection across the *Mus* genus

Species: *M.spicilegus, M.macedonicus, M.pahari, M.caroli, M.spretus, M.m.musculus and M.m.domesticus*

	M1 v M2 (p-value)	M7 v M8 (p-value)	Number of codons analyzed	Number of of positive sites	Percent positive sites
HJURP	0.0061658	0.004768	654	22	3.4
CENPA	0.7587	0.4992	132	5	3.8
CENPB	1	1	456	2	0.4
CENPC	1.15E-07	1.90E-08	870	32	3.7
CENPT	0.9737	0.9458	500	11	2.2
Mis12	1	1	207	0	0.0
Mis18a	1	1	203	0	0.0
Mis18bp1	1	1	847	0	0.0

Figure 3: Centromeric protein-coding genes are evolving differently across different branches of the Mus genus. A. Evolutionary changes in *M. spicilegus* and *M. macedonicus* centromeric protein-coding genes are concentrated at specific regions when compared to the reference *M. spretus* genes. These identified domains may be categorized by specific function, with some domains possibly relevant to centromeric binding and targeting. These domains would provide a stronger link between evolution and centromeric protein function, especially if changes in critical domains are identified between multiple species, which would demonstrate a strong evolutionary trend. **B.** Short-term positive selection using pairwise analysis of *Mus* species pairs. dN/dS is the ratio of the rate of nonsynonymous amino acid (dN) and the rate of synonymous amino acid changes (dS) between two species that determines the nature of ongoing selection pressure. dN/dS = 1, neutral selection, dN/dS < 1, negative selection, dN/dS > 1, positive selection. A dN/dS value ≥ 1, indicating neutral or positive selection is highlighted in red. NA indicates a value that cannot be calculated due to either no synonymous (dS) or nonsynonymous (dN) changes

between the two compared sequences. **C.** Long-term positive selection analysis across several *Mus* species using Codeml analysis. Likelihood of models of neutral codon evolution (M1 or M7) were compared to models allowing positive selection (M2 or M8). Codons with insertions, deletions, and ambiguous alignments were excluded from the analysis. The number of positive selection sites is the number of codons with p > 0.95 Bayes Empirical Bayes (BEB) analysis from M8. A Likelihood Ratio Test (LRT) p-value <0.5 indicating positive selection is highlighted in red.

2. Centromeric proteins exhibit differences in centromere localization in early and late diverging *Mus* species.

Following the characterization of the pattern of evolutionary changes in centromeric protein-coding genes, I investigated how the function of these centromeric proteins has changed during evolution. Localization of centromeric proteins to centromere determines their function in chromosome segregation. Using cytological visualization, I investigated how localization of these proteins to centromeres is different in *Mus* species. I first developed a protocol to prepare immunostaining grade mouse liver cell spreads. I tested several methods, including hypotonic treatment followed by fixation and dropping cells on the slide and a direct tissue spread method followed by fixation. The direct tissue spreading protocol produced high-quality spreads comparable to those produced from a cell line (Figure 4). Next, I standardized antibodies against centromeric proteins in 3T3 cell lines. CENP-A alone was stained with anti-CENP-A antibodies and several centromeric proteins were stained together with CREST antibodies (Supplementary Figure 1). CREST serum is derived from CREST (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia) patients and contains antibodies against several centromeric proteins, including CENP-A, CENP-B and CENP-C (Mayo Clinic, 2023).

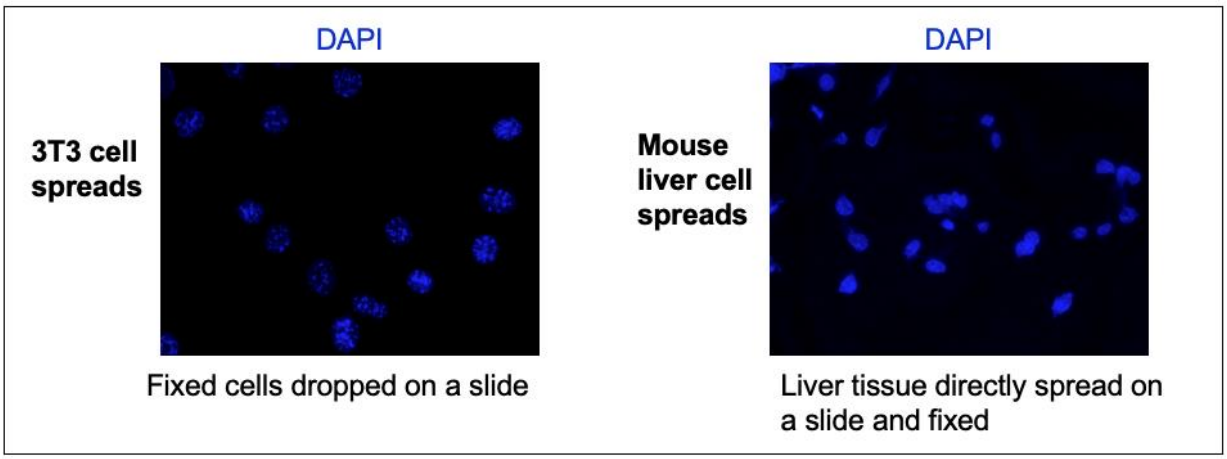


Figure 4. Preparation of cell spreads from mouse liver tissues. Direct spreading of liver tissue on slides results in high quality cell spreads comparable to those produced from a cell line. DAPI stained nuclei from 3T3 cell spreads using dropping method (Left) and mouse liver cell spread produced using the direct spreading method (Right).

Following the standardization of liver cell spreading protocol and antibody staining, I stained centromeric proteins in the liver cell spreads from various early and late diverging *Mus* species (Figure 5A). To determine the relative levels of centromeric proteins at centromeres, total cell fluorescent counts were analyzed for CREST and CENP-A (Figure 5B). I found that the relative intensity of both CENP-A and CREST staining differed in *Mus* species analyzed. These findings reveal specific functional differences at centromeres in the form of centromeric protein localization in early and late diverging *Mus* species and require further investigation, including replicates. These functional differences might have resulted from both early short-term and long-term evolutionary changes in the *Mus* genus.

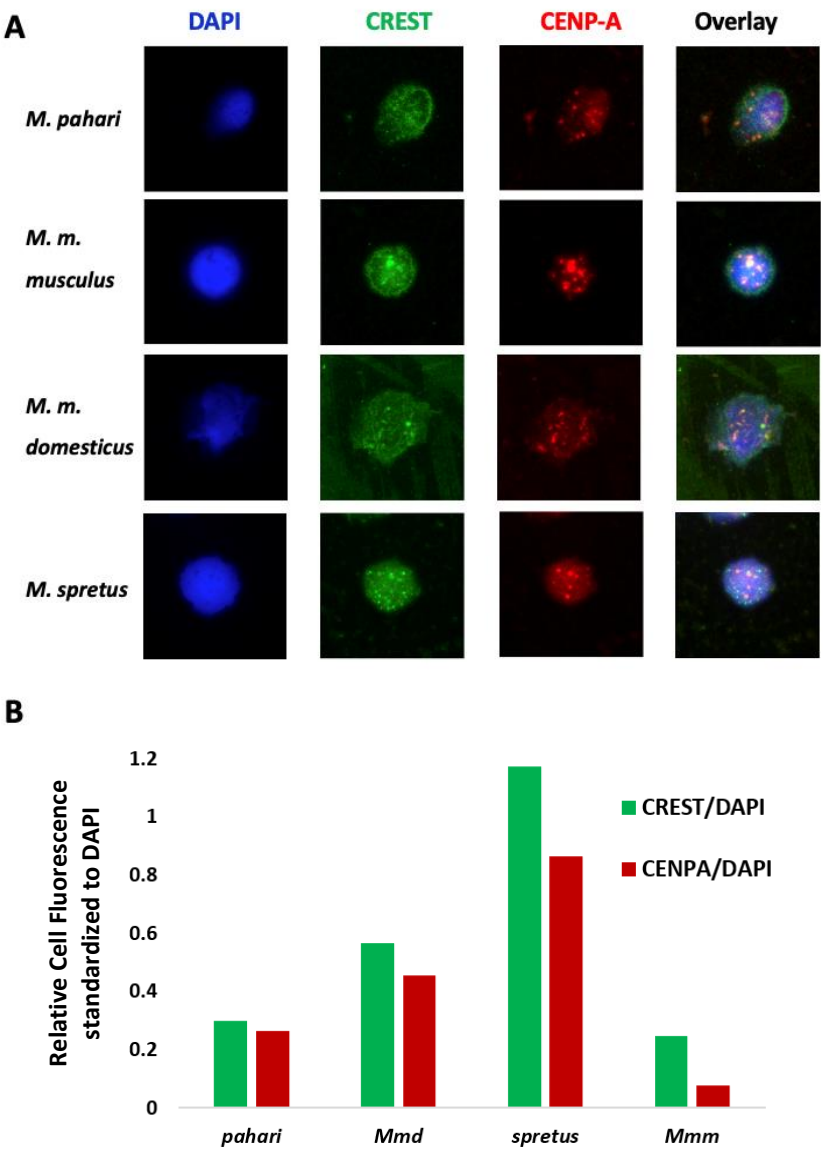


Figure 5. Differences in the localization of centromeric proteins in *Mus* species. A. Images showing DAPI, CENP-A and CREST immunostaining in *M. Pahari*, *M. m. musculus*, *M. m. domesticus* and *M spretus.* **B.** Normalized relative cell fluorescence calculated by dividing total cell fluorescence of CREST and CENP-A, by the fluorescence of DAPI for the respective species.

3. Developing a cytological approach to determine copy number variations for centromeric protein-coding genes.

In addition to nucleotide level changes, genomic loci can undergo copy number variations (CNVs) during evolution. CNVs are particularly interesting for rapidly evolving essential genes as the multiple gene copies provide additional opportunities for gene expansion and differentiation. I

hypothesize that CNV contributes to the rapid evolution of centromeric protein-coding genes in the *Mus* genus. CNVs are generally detected by computational methods that rely on the sequencing depth of the Whole Genome datasets and can report low confidence CNVs in datasets with a low sequencing depth, which is often the case with datasets from non-model organisms such as *Mus* species other than the reference *M. musculus* strain. Therefore, I decided to utilize a more direct and deterministic FISH (Fluorescent in-situ hybridization)-based Oligopaint method to determine CNVs by visualizing single-copy genomic loci (Beliveau et al. 2014). The Oligopaint-FISH bypasses the need to utilize bacterial artificial chromosome (BAC) probes, which are difficult to obtain. Instead, Oligopaint-FISH uses several short oligos spread across the genomic locus of interest, making it a versatile method for visualizing any genomic locus (Figure 6). Since HJURP emerged as one of the centromeric protein-coding genes that exhibited long-term adaptive evolution in the *Mus* genus, CNVs in *Mus* species might be responsible for HJURP evolution. Therefore, I decided to standardize HJURP Oligopaint. First, we constructed an oligo library (Supplementary Table 3) spanning the HJURP locus. Next, the library was amplified using fluorescently labeled primers and nicked to generate a single-stranded probe, which was then gel extracted and purified (detailed protocol in the methods section). The purified probe was used to visualize HJURP loci in 3T3 cells. HJURP Oligopaint probe produced four separate dot-like signals in 3T3 cells known to be tetraploids. These results demonstrated the success of the Oligopaint method for detecting HJURP gene copy numbers (Figure 7). In the near future, this HJURP Oligopaint probe will be used to determine the number of HJURP copies in the liver cell spreads from various *Mus* species.

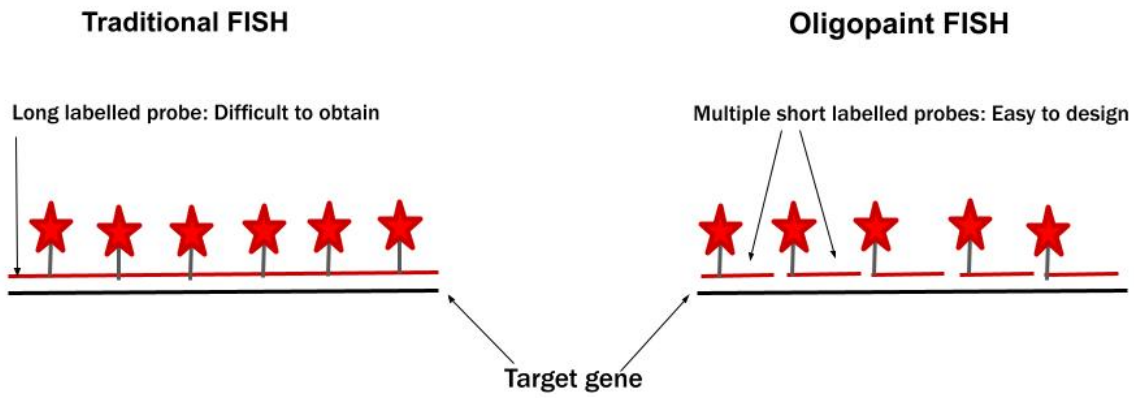


Figure 6. Oligopaint-FISH is a superior cytological visualization method compared to the traditional FISH bioinformatic approach. Traditional FISH involves using long labeled BACs as probes that are much more difficult to obtain than the short, labeled oligos used as probes in Oligopaint-FISH. Additionally, long labeled BAC probes bind to large genomic regions leading to non-specific signals, which can pose a serious problem while studying target genes undergoing evolutionary change. Oligopaint FISH probes bind to precisely defined sequences leading to the desired amount of fluorescence and specificity.

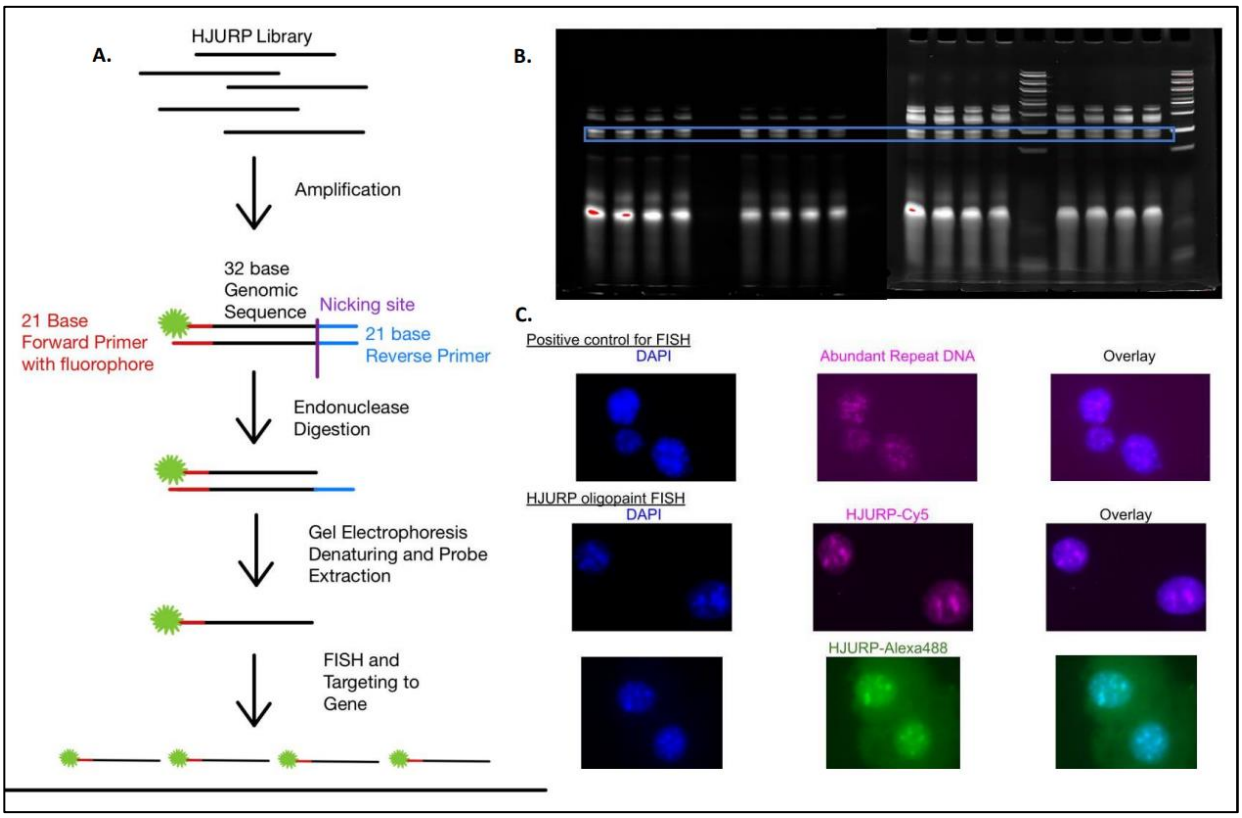


Figure 7. Standardization of Oligopaint FISH for HJURP gene in a cell line. A. A schematic of HJURP Oligopaint probe preparation. Two different fluorophores for the forward primer, Cy5 and Alexa488 were tested. The Alexa488 probe produced brighter signals as compared to the Cy5 probe. **B.** The 53 bp probe DNA band on TBE-Urea gel marked with the blue box was extracted and purified (Left-Alexa488 channel, right - ethidium bromide). **C.** Purified probe was used to perform FISH in 3T3 cells. Abundant repeat DNA was used as a positive control for the FISH experiment.

Discussion

Findings present in this study suggest that the pattern of selection pressure on centromeric protein-coding genes varies in different branches of the *Mus* phylogenetic group. While HJURP and CENP-C are subjected to long-term adaptive evolution, CENP-T shows only short-term

positive selection in selected phylogenetic branches. *M. spicilegus* and *M. macedoncius* display a unique pattern of centromeric protein-coding genes in the form of short-term selection on CENP-T. Furthermore, the long-term adaptive evolution of HJURP and CENP-C does not result from a common evolutionary selection pressure on all the *M*us species. Future studies will focus on further investigating the differences in the localization of centromeric proteins in both early and late diverging lineages/species in the *Mus* genus, including replicating the results presented in Figure 6, as well as determining centromere localization of additional centromeric proteins for which the antibodies have been standardized (Supplementary Figure 1) in these *Mus* species. Additionally, several other *Mus* species available in the lab will be analyzed to determine if these functional differences in centromeric protein localization are common to the entire *Mus* genus. Furthermore, *M. macedonicus* mice will be obtained to determine if CENP-T, which shows short-term rapid evolution in these species, exhibits functional differences.

In addition to nucleotide changes, CNVs can significantly impact the evolution, especially the adaptive evolution, of essential genes. I have standardized the Oligopaint-FISH method for determining the HJURP copy number. In the near future, HJURP Oligopaint FISH will be used to determine HJURP CNVs in several species from the *Mus* genus. Furthermore, the Oligopaint method can be customized for other centromeric proteins of interest, such as CENP-A, by generating a new library pool to determine such CENP-A CNVs across species. Ultimately, this evolutionary work contributes toward our understanding how evolutionary changes occur and how they impact centromere function, which will provide insights into broader mammalian cell division function and organismal health.

Methods

Assembly, alignment, and analysis of centromere protein-coding gene transcripts

Transcriptome data for *Mus spicilegus* was downloaded from DRA search, and RNA sequence data for *M. macedonicus* was downloaded from NCBI SRA. Protein coding sequences for *M. Spretus* were downloaded from Ensembl. Protein coding genes analyzed included HJURP, CENP-A, CENP-B, CENP-C, CENP-T, Mis12, Mis18a, Mis18b, and Mis18bp1. Reference sequences (*M. spretus* or *M. spicilegus*) were utilized to generate a BLAST (Basic Local Alignment Search Tool) analysis of transcriptome data. BLAST sequences were aligned in DNADynamo and adjusted for accuracy. DNADynamo generated and exported consensus sequences for protein-coding genes. Mutations between consensus sequences were displayed in the DNADynamo program.

Pairwise Analysis of *Mus* Species

Protein coding sequences for *M. caroli*, WSB, PWK, and *M. pahari* were downloaded from Ensembl. Consensus protein coding sequences of seven species (WSB, PWK, *M. caroli, M. pahari, M. spretus M. spicilegus, and M. macedonicus*) were aligned in DNA Dynamo, where gaps and affected codons were removed, creating an accurate primary alignment for all species. If a majority of species aligned well, but one or two species caused significant disruptions to the alignment, the excepting species were removed to create a secondary alignment with fewer codons removed. The secondary alignment was used to produce more accurate ratios between species, with the primary alignment filling in the gaps. The primary alignment was converted to PAML format and used to create a maximum relatedness evolutionary tree. This tree and the sequences were applied to CodeML, which identified sites of positive selection. Finally, the

primary and secondary alignments were run in SNAP, Synonymous Nonsynonymous Alignment Program, to generate dN/dS ratios between species. (Korber, 2000 and www.hiv.lanl.gov/).

Oligopaint HJURP Probe Preparation

HJURP Probe amplification primers

Custom 5'O F Primer: TGAATAGCAGCGGTGGCAAAC

Custom 3'O F primer: (containing Nb.BsrDI site): GTATCGTGCAAGGGTGCAATG

HJURP Oligopaint library. HJURP probe library sequences are given in Supplementary Table 3. An example probe from the library is given below.

TGAATAGCAGCGGTGGCAAACTTTAATTTATTAGCAGTTTGCTGTCAGCAGCTTGGCTGGTTT**CATTGC**A CCCTTGCACGATAC

Amplification and Precipitation. The 100 ul PCR reactions were set using 3.76 ul Forward primer with a fluorophore, 1.06 ul reverse primer, 1 ul HJURP library (0.44 pg/ul), 45.18 ul dH2O, and 50 ul EmeraldAmp® GT PCR Master Mix. The reactions were then set in a thermocycler for 40 cycles with the following parameters: 98 ℃ for 10 seconds, 50 ℃ for 30 seconds, 72 ℃ for fifteen seconds, with cycles followed by 72℃ for five minutes. The reactions were stored at 4 degrees between use. To precipitate the PCR products, 1-2 ul glycogen, 80 ul 2M ammonium acetate, and 1 mL of ice-cold 100% ethanol were added to 400 ul of PCR reaction, mixed thoroughly, and incubated at -80°C for 1 hr or at -20°C for at least 3 hours. After incubation, the mixture was centrifuged for 35 minutes at 4 °C and 13,000 rpm. The supernatant was decanted, and 1ml 70% ethanol was added to each tube. The mixture was centrifuged again for 20 min at 4 °C and 13,000 rpm. The supernatant was decanted, and the pellet was dried. The pellet was then resuspended in 40 ul ddH20.

Nicking Endonuclease Digestion and Concentration. A reaction mixture containing 20 ul amplified DNA, 10ul Nb.BsrDI Digestion enzyme, 10 ul Cut Smart Buffer, and 60 ul ddH20 was prepared. Incubated the tubes at the following program in a thermocycler: 65°C for 4 hours, 80°C for 20 minutes, 4°C for one hour. Precipitated as described previously, with the final pellet resuspended in 60 ul ddH2O. Incubated at 37°C for 30-60 minutes to ensure proper DNA suspension.

Electrophoresis, Gel Extraction, and Recovery. Preparing 15% TBE-UREA polyacrylamide gel: Make 10 ml TBE-UREA-PAGE solution containing: 5 ml 30% acrylamide solution, 1 ml 10X TBE, and 4.8g UREA. Added dH2O to make the final volume 10 ml. Mixed the solution thoroughly and then added the following: 100 ul of 10% APS, 10 ul of TEMED (flammable). A 10 mL solution is sufficient for 2 gels. Mixed thoroughly after addition. The gel starts polymerizing as soon as APS and TEMED are added, so those reagents were added last. Poured the solution into gel plates without waiting. The solution was poured, the comb was inserted, and the gel was allowed to solidify.

Setting up the electrophoresis system: Heated 1L of 1X TBE buffer for approximately 30 mins at 60-65 °C to prevent renaturation. Ensured buffer was not boiling to avoid any crack. Filled the tank with the hot 1X TBE buffer. Ran the gel (preloading) at a constant 20 W.

Sample preparation and running: Added 15 ul 2X TBE-Urea sample buffer to 15 ul pooled digestion product or 15 ul DNA ladder. Vortexed and incubated at 95°C for 5 minutes in a thermocycler, then kept on ice. Loaded the samples into the gel. Ran the gel for approximately 1 hour at a constant 100 W, or until the blue sample buffer reached the bottom of the gel.

Staining and extracting the ssDNA: Removed the gel from the plastic casing and stained with 10 ul ethidium bromide for 5 min, de-stained with ddH2O for 5 min and visualized the gel on a UV

box. Stain and de-stain steps occur in sufficient 1x TBE buffer such that the gel is completely covered and can agitate. Imaged gel in Ethidium Bromide channel and a channel appropriate for the fluorophore to help determine the necessary band. Cut out the approximately 53 bp band using a razor blade (Figure 5c). Added each gel slice to a 2 ml Eppendorf containing 3x the mass volume of Diffusion buffer and crushed the gel. Diffusion buffer: 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% SDS at pH 8.0. Incubated overnight at 55°C. The next day, transferred the liquid from each tube to a fresh 2 ml Eppendorf using a syringe. Added 1-2 ul glycogen and 3 times the volume of ice-cold 100% ethanol to each tube, vortexed vigorously and precipitated as described before. Added 10 µl of ddH2O to each pellet. Incubated at 37°C for 30-120 minutes to ensure DNA suspension. Quantified DNA using a tapestation or Nanodrop.

Liver Tissue Cell Spread Production

Direct Spreading of Snap Frozen Tissue: Frozen mouse liver tissue was removed from -80°C, and a small piece, approximately 0.5 grams, was extracted. The tissue extract was washed twice in 1X PBS. After washing, a portion of the tissue, approximately 0.1 grams, was placed on a standard microscopy slide with 60 ul of 1X PBS. Using curved forceps, tissue was finely minced until the PBS liquid became cloudy. Additional PBS was added until liquid was evenly distributed on the slide. A standard glass coverslip was gently placed on the slide, and the slide was placed in liquid nitrogen for approximately one minute. The slide was removed from the liquid nitrogen and the coverslip was removed with a fresh scalpel. The slide was immersed in a chilled jar containing acetone for three minutes.

Homogenized Cell Spread Method: Obtained homogenized frozen mouse liver cells from -80°C, approximately 10 million cells in a 1.5 ml sample. Defrosted and washed cells in 5 ml PBS. Added

5 ml of prewarmed to 37C 75 mM KCL dropwise while gently shaking. Incubated cell solution for 15 minutes at 37ºC. Freshly prepared sufficient methanol: glacial acetic acid (3:1) fixative solution. Added 5 drops of freshly prepared fixative, invert, and spun cell solution at 2000 rpm for 10 minutes. Decanted supernatant leaving 250 ul of hypotonic solution. Resuspended cell pellet by tapping and added 5 ml of fixative dropwise for the first 1 ml while gently shaking. Incubated for 20 min at 4ºC. Spun down cells at 2000 rpm for 10 min, and decanted leaving 250 ul of solution. Resuspended cells, and continued the process of adding 5ml of fixative, incubating, and spinning down cells approximately 3-6 times, with subsequent incubation times lasting 10 minutes, until cell pellet was colorless and sufficiently washed. Resuspended the pellet in 500 ul of fixative to create a cloudy cell solution. Drew the cell solution into a glass dropper and from a height greater than four feet, dropped the cells onto standard microscopy slides. Allowed the slides to air dry for three hours before use.

Oligopaint FISH

Obtained microscopy slides with fixed cells and ensured slides were at room temperature. Incubated slides in 2X SSCT + 50% (v/v) formamide for 5 minutes at 82°C in a prewarmed coplin jar. Removed lids from jars to allow them to reach temperature. Checked that solution was at the appropriate temperature with a thermometer. Transferred slides to a coplin jar containing 2X SSCT + 50% formamide at 60°C and incubated for 20’. Removed the slides from the 60°C coplin jar and allowed them to cool to room temperature. Created 25 ul Hybridization cocktail (per slide) in thin-walled PCR tubes. The hybridization cocktail consists of 2x SSCT (0.1% Tween) 50% formamide solution with 10% Dextran Sulfate, containing 10-20 pmole of Oligopaint Probe DNA, and 1 uL RNase A. Cut the bottom of PCR tubes and squeezed 25 µl of a hybridization cocktail to

a 22x22 coverslip. Inverted slides onto the cocktail-containing coverslips such that the area containing the cells is covered. Sealed with rubber cement. Allowed the rubber cement to air-dry for 5' at room temperature. Denatured for 5' at 82°C by placing slides top of a water-immersed heat block inside a water bath or using the thermocycler. Transferred slides to a humidified chamber and hybridized overnight at 37°C. The next day, removed the coverslip carefully and washed slides in a pre-warmed coplin jar containing 2X SSCT at 60°C for 15 minutes. Transferred slides to a coplin jar containing 2X SSCT at RT and incubated for 10 minutes. Transferred slides to a coplin jar containing 0.2X SSC at RT and incubated for 10 minutes. Removed slides from the coplin jar, remove extra liquid, and mounted with 2X DAPI and applied a coverslip. Sealed each slide using nail polish and ensured dry before imaging.

Immunostaining and Analysis

Seeded cells in 4-well slide at 150,000-300,000 cells a well. Let grow overnight. Checked cells for confluency, around 70% is ideal. Fixed cells with acetone by adding 250 ul of cold acetone to each well. Incubated for 5 minutes at -20°C. Fixed cell spreads can serve as a substitute for seeded slides. Discarded acetone, and incubated in 1% BSA 1X PBST (0.1% Triton X) for 30 minutes at room temperature. For all subsequent steps, used 250 ul solution per well for seeded slides, and 500 ul solution per fixed cell spread slide. Discarded liquid, and incubated for at least three hours at 4°C in 1% BSA 1X PBST containing primary antibody. Discarded antibody solution, and washed slides 3x for 5 minutes in appropriate volumes of PBST. Incubated slides with secondary antibody in PBST for one hour at room temperature. Discarded solution, and washed slides 3x for 5 minutes in PBST. Mounted slides with 2X DAPI solution and a coverslip. Sealed each slide using

nail polish and ensure dry before imaging. Slides remain stable for several months when stored at 4°C in dark.

Fluorescence intensity analyses

Using ImageJ, converted images to 16-bit grayscale and selected threshold intensity range from DAPI stained image. Utilized this threshold image to analyze other particles in other channels representing antibodies. Loaded fluorescent antibody image and outlined a cell of interest with a selection tool. Measured area integrated intensity and mean grey value. Selected an area of outside of the cell with no fluorescence for the background and took measurements. Used the following formula to find CTCF (Corrected Total Cell Fluorescence): CTCF = Integrated Density – (Area of selected cell X Average fluorescence of background). CTCF values were then normalized to DAPI.

References

Beliveau, B. J., Apostolopoulos, N., & Wu, C. T. (2014). Visualizing genomes with Oligopaint FISH probes. Current protocols in molecular biology, 105, 14.23.1–14.23..

https://doi.org/10.1002/0471142727.mb1423s105

Bloom, K., Costanzo, V. (2017). Centromere Structure and Function. In: Black, B. (eds) Centromeres and Kinetochores. Progress in Molecular and Subcellular Biology, vol 56. Springer, Cham. https://doi.org/10.1007/978-3-319-58592-5_21

Booker, T. R., Jackson, B. C., & Keightley, P. D. (2017). Detecting positive selection in the genome. BMC biology, 15(1), 98. https://doi.org/10.1186/s12915-017-0434-y

Couger M. B., L. Arévalo, and P. Campbell, 2018 A High Quality Genome for *Mus spicilegus*, a Close Relative of House Mice with Unique Social and Ecological Adaptations. G3: Genes, Genomes, Genetics 8: 2145–2152. https://doi.org/10.1534/g3.118.200318

Dunleavy E. M., D. Roche, H. Tagami, N. Lacoste, D. Ray-Gallet, et al., 2009 HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell 137: 485–497. https://doi.org/10.1016/j.cell.2009.02.040

Dunleavy, E. M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., & Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell*, *137*(3), 485–497. https://doi.org/10.1016/j.cell.2009.02.040

Harris, E. E., & Meyer, D. (2006). The molecular signature of selection underlying human adaptations. *American journal of physical anthropology*, *Suppl 43*, 89–130. https://doi.org/10.1002/ajpa.20518

Kodama, Y., Shumway, M., Leinonen, R., & International Nucleotide Sequence Database Collaboration (2012). The Sequence Read Archive: explosive growth of sequencing data. *Nucleic acids research*, *40*(Database issue), D54–D56. https://doi.org/10.1093/nar/gkr854 , used in conjunction with DDBJ (nig.ac.jp)

Korber B. (2000). HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences, Chapter 4, pages 55-72. Allen G. Rodrigo and Gerald H. Learn, eds. Dordrecht, Netherlands: Kluwer Academic Publishers. Used alongside www.hiv.lanl.gov/ .

Malik H. S. (2009). The centromere-drive hypothesis: a simple basis for centromere complexity. Progress in molecular and subcellular biology, 48, 33–52. https://doi.org/10.1007/978-3-642-00182-6_2

Kumon, T., Ma, J., Akins, R. B., Stefanik, D., Nordgren, C. E., Kim, J., Levine, M. T., & Lampson, M. A. (2021). Parallel pathways for recruiting effector proteins determine centromere drive and suppression. Cell, 184(19), 4904–4918.e11. https://doi.org/10.1016/j.cell.2021.07.037

Mayo Clinic Laboratories Neurology Catalog. (2023). Centromere Antibodies, IgG, Serum - Mayo Clinic Laboratories | Neurology Catalog. Retrieved March 20, 2023, from https://neurology.testcatalog.org/show/CMA

Miller, C. H., Campbell, P., & Sheehan, M. J. (2020). Distinct evolutionary trajectories of V1R clades across mouse species. *BMC evolutionary biology*, *20*(1), 99. https://doi.org/10.1186/s12862-020-01662-z

Morrison, O. (2021). Rapid Adaptive Evolution of Centromeric Proteins in the Mus Genus. Thakur Lab presentation, unpublished.

Perpelescu, M., & Fukagawa, T. (2011). The ABCs of CENPs. Chromosoma, 120(5), 425–446. https://doi.org/10.1007/s00412-011-0330-0

Sanchez-Pulido, L., Pidoux, A. L., Ponting, C. P., & Allshire, R. C. (2009). Common ancestry of the CENP-A chaperones Scm3 and HJURP. Cell, 137(7), 1173–1174. https://doi.org/10.1016/j.cell.2009.06.010

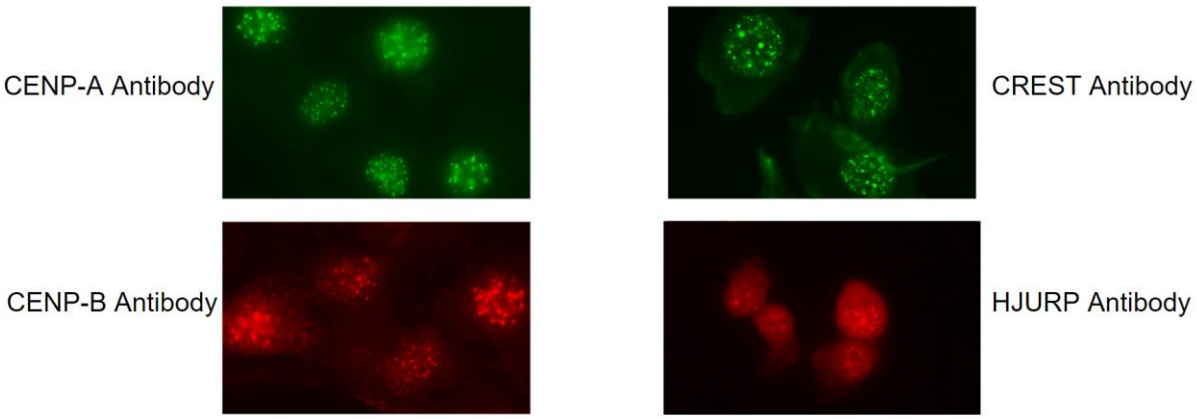
Sequence Read Archive (SRA) [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2009 - [cited 2023 03 20]. SRX7400223: RNA-

Seq of Mus macedonicus: pooled male and female adult vomeronasal organ. Data originally published in Miller et al. 2020. Available from: https://www.ncbi.nlm.nih.gov/sra/

Supplementary Information

Standardization of antibodies against centromeric proteins for immunostaining.

To demonstrate the feasibility of staining various centromeric proteins, antibodies for the centromeric proteins HJURP, CENP-A, CENP-B, and CREST, an anticentromere antibody, have been standardized in the mouse 3T3 cell line (Figure S1).



Supplementary Figure 1. Standardization of antibodies against centromeric proteins in **3T3 Cell line**. CREST antibody - 1:500 dilution, Anti-CENPA antibody - 1:500, Anti-CENPB antibody - 1:1000, Anti-HJURP antibody -1:250.

Supplementary Table 1. *M. spicilegus* constructed centromeric protein coding gene transcripts

Name	Sequence
CENP-A	ATGGGCCCCCGTCGCAAACCGGAGACCCCAAGGAGGAGACCCTCCAGCCCGGCGCCTGGACCCTCGCGACAGAGCTCCAGTGTAGGCTCTC AGACACTGCGCAGAAGACAGAAATTCATGTGGCTTAAGGAAATCAAGACCCTGCAGAAGAGCACAGACCTCTTGTTCAGGAAGAAGCCTTT CAGCATGGTTGTTAGAGAAATATGTGGGAAGTTCAGCCGTGGTGTGGATTTATGGTGGCAAGCCCATGCCTTGTTGGCCCTTCAGGAGGCA GCAGAAGCTTTCCTCGTCCACCTCTTTGAGGACGCCTACCTCCTCTCCTTACATGCTGGTCGGGTCACGCTTTTCCCCAAAGACATTCAGTTGA CCAGGAGAATCCGAGGCTTCGAGGGCGGACTCCCCTAA
CENPB	ATGGGCCCCAAGCGGCGGCAGCTGACGTTCCGGGAGAAGTCGCGGATCATCCAGGAGGTGGAGGAGAACCCGGACCTACGCAAGGGCGA GATCGCGCGGCGCTTCAACATCCCGCCGTCCACGCTGAGCACCATCCTGAAGAACAAGCGCGCCATCCTGGCGTCGGAGCGCAAGTACGGA

	GTGGCCTCCACCTGCCGCAAGACCAACAAGCTGTCCCCGTACGACAAGCTGGAGGGGCTTCTCATCGCTTGGTTCCAGCAGATCCGCGCCG CCGGGCTGCCGGTCAAGGGCATCATCCTGAAAGAGAAGGCGCTGCGGATAGCGGAGGAGCTGGGCATGGACGACTTCACTGCTTCCAACG GCTGGCTGGATCGCTTCCGCAGGCGCCACGGTGTAGTGGCCTGCAGCGGCGTGACCCGTTCTCGGGCGCGAAGCTCTGCCCCTCGGGCCCC AGCGGCACCTGCCGGCCCAGCCACGGTGCCCTCCGAGGGCAGCGGTGGAAGCACACCGGGCTGGCGCACGAGGGAGGAGCAGCCGCCGT CGGTAGCTGAGGGCTACGCCTCGCAGGACGTGTTCAGCGCCACCGAGACCAGCCTGTGGTACGACTTTCTGTCGGACCAGGCCTCAGGGCT CTGGGGAGGTGATGGACCGGCGCGCCAGGCCACCCAGCGCCTTAGCGTTTTGCTGTGCGCCAACGCCGATGGCAGCGAGAAGCTTCCCCC ACTGGTCGCAGGCAAATCTGCCAAGCCCCGTGCAGGGCAAGGTGGTCTGCCCTGCGACTACACTGCCAACTCGAAGGGTGGAGTCACCACC CAGGCTCTGGCTAAGTACTTGAAAGCTCTGGACACTCGAATGGCAGCAGAATCTCGTCGGGTCCTTCTGCTGGCCGGCCGTCTGGCTGCCCA GTCCTTGGACACCTCGGGCCTGCGGCACGTGCAGCTGGCCTTCTTCCCTCCTGGCACCGTGCATCCTTTGGAGCGAGGAGTGGTCCAACAGG TGAAGGGCCACTACCGCCAGGCGATGTTGCTCAAGGCCATGGCAGCACTCGAGGGTCAGGATCCCTCAGGCTTGCAACTGGGCCTTGTGGA GGCCCTGCACTTTGTGGCTGCAGCCTGGCAGGCAGTGGAGCCCTCGGACATAGCAACTTGCTTTCGTGAGGCGGGTTTTGGAGGTGGTCTT AATGCCACTATCACCACTTCCTTCAAGAGTGAGGGAGAGGAAGAGGAGGAAGAGGAGGAGGAAGAGGAGGAGGAGGAAGGGAGGGTG AAGGGGAAGAGGAAGAGGAAGAAGAAGGGGAGGAGGAAGAAGGAGGGGAAGGAGAGGAGGTAGGAGAGGAAGAGGAAGAGGTAGA AGAGGAGGGTGAGGTTGATGATAGTGATGAAGAGGAGGAAGAAAGTTCCTCTGAGGGTTTAGAGGCTGAAGACTGGGCCCAGGGAGTA GTGGAAGCCAGTGGTGGTTTTGGGGGTTATAGTGTCCAGGAAGAGGCCCAGTGCCCCACTCTCCATTTCCTGGAAGGAGGAGAGGACTCT GACTCAGACAGTGATGAAGAGGAGGACGATGAAGAGGAGGATGAGGATGATGATGATGAGGATGATGATGAGGATGGTGATGAGGTCC CTGTGCCCAGCTTTGGGGAGGCCATGGCTTACTTTGCCATGGTCAAGAGGTACCTGACCTCCTTCCCCATCGATGACCGAGTGCAGAGTCAC ATCCTTCATTTGGAACACGATCTGGTCCACGTGACTAGGAAGAACCATGCCAGGCAGGCGGGAGTTCGGGGTCTTGGACATCAAAGCTGA
CENP-C	GGGACGCCTCCGTCCGCAATCTTACTCTAGTCCTCCTGGAACCCTGGCCGGCTTGGGTTTCCCTCTCAGGCCTCCGTTTGACCTAGTGGGCCG GAACCATGGCCTCGTTCCATCTGGATCATCTAAAGAACTACCACAGAAGATATTGCCGGCCTTCCAGAGCACCCAACATTCATACAAAAAAA GGCCAGAACATGCTGGAAATCTTACAAGATTGTTTTGAAGATCAGAGTAAGGCTTCATTTCTTGATGATTTTACAGAGTCATTGACTTCTTCA ACACAAAAAAAGAAAGCTAATTATAGTCAATCATCAAGCAAAAAGTGCCCAGAGTCACATTCCAAACCAGTTCCAGTGTCTTCAAGGACGG GAGAAGCTTCTTTGCAAGCGTCTGCAGAGCCGAGTGAAGCCGCTGGTGGATCAGTTCAGGCTAATGAAGTTCACCATGGTGTGTCTGATGA ATTGGACTTATGTGTTCGCTCCCCTGTGGTTCTCCTGGATGCAAATGTCAACACGTTACAGAAAGCCGCTTCACCTGCTGGCCAAAAGAGAG TGGCTTCTGTCTCAAGGAGTCCAGTAGATAGGCAGGCTTCAAATAAAAACATTTCTTTTAAAACCAGGAAAAGGTTAAACTTTGAAGATAAA GTTACTTCGAGCACAGCAGAAACAGAGAACAGTGTGTTACAAGTAGAGGATAACTTATCCAAGGGACAAGAAGGGACATCATCAGAAATC ACTCAGAAAAGAGACGATCTGAGTTCTGATGTCCAATCACGTTCTAAGAAGAATTTTTCGGAGTTGTTTTTAGAGACTGTCAAGAGGAAAAG TAAATCCAGTTCTGTGGTCAGGCACACAGCAGCAGTTCCATCTTCTCCGCCTCCTCCAAGTGACATGAAATTATTAGAAGATGAATTTATTAT TGATCGGTCAGATAGAAGTTTTTCAAGTCGACTTTGGGTGATGATACCAAGCAAAGACAGACATCTGAGCGCCCACAAGCCATCTCCTGAAA ACACTGCACTGCTTCAAGGTAAGAAGTCAAGAGAGAAATCTCATAGTTTATCTGCAATGACGTTTGCAAGAAACACACAGTCTGATAAGGCT CACCCAATAGAGGAAGCTCAGCTCTCTGTTGAAGAAAACCTGGGAACTACTTGTACAGATGAGTTGGAAAATGATTGTCGATCTCCTGAAAA TAAAATGCAATCTGAGACTGCCAAAACCCCACCCGCATGGGAAAGGACTACGAAACAGAACCAGAGGAGAGTGTCTAAGCCTAAAGCAGC TGAAGAGCTCAGGAAGGGACAGAGTAGCTGTGAAAATAGCAATGTGTCAAACACTGGCCAAGACAAGCTACAGATCAACTCGAAAAGAAA TATGAAGGACTGTGAAGAGGTGAGGAATGAGCCTACTCCCAAGAAACAGAAGCCAGCTCTTGAAAATAAGAAAAAAAACAATAGCACCCA AACCAATAAGGAAAAATCTGGAAAGAAATTTTTTTCTGGTGGGTCCAAGAACAAATTTGTTCCCAAAAAAGTAACTTTGACTTCCAGGAGAA GTTGTAGAATTTCTCAGCGTCCATCTGACTGGTGGAGGGTAAAATCAGATGAAAGTTCTGTTGATAGAAATCCTTCAAAAGAAAATAACTCA CCAGTGGTATATCCCAATAAAAAGAAGCAGACTAAGAGAAATCATGTATCTAAGAGGGCTGGGAAGAAACCTGGTCCATCAAAAAGACAG AAGACAGAGATGAGCCCAAGAGTACAGAAGTCTTTAAATGTAAAGGATTCTGGAGGAACTGTTAGTGGCCATGATGACACTTCTCGTTCCC AGAGGAAGCCATTGAAAAACATTGAGGCAGACCCAACTCAGAAGAGCCTTGCTATTTCTCGACCTAAAAGAGGCTGTAAGTATCGAAACAA TGTTATGACATCACCGAATGTTCATTTAAAGTCTCATACTGAGGAATATACAAGCAAGACACAAATGGAGTCTGCTTCAAATTCAGAAGTGT CTAAACGTTCAGTGTGGGAAGAAAGTGGACCTTCCAGGTTCAAGAATTATGAAATGCCAGGGAGCAGTAATTCTGAGATGGGTGATGAAC AGGACCAGAAGAGTTTGCATTTTACAACAAGAAGTTTTAATATGGTACCAGATAAAAAATTACATCACAAATTAGTACTGCCCTCCAACTCAC CAAATGTTCGTCGATCTAATAGAATACGTTTGAAACCTCTGGAGTATTGGCGAGGGGAACGGGTAGATTATCAAGAAAGTTCGTCGGGACA ACTTGTGCTTGAAATAATATCCCCAGCTTCAGTACCCACGAAAATAAAGGCACAAAGAAACCTCGGCAAAGTCAACAAAAAAGTTACCAAGA AGCAGACCCATCTCAATAGCCATGAAAAGGCTAAGATGGAGCTGCCTCTGGATATGCATCTGGGAGATCCCTTTCAGGCAACATTGGCGAA GGACCCAGAAACCGCAGAGCTTGTCCCCATGGATCTTATAAGTCCACGAGATACACACCGCTTTTTTGTTGAGCAACATGGTTTGAAAGTTTT TAAAACATTGGATACAATATATTTTTCTACTGGAAAATTGGTTTTAGGACCCTATGAAGAGAAAGGAAAACAGCATGTTGGCCAAGATATAT TGGTTTTTTATGTTAACTTTGGTGACCTTTTATGTACATTACATGAAACACCTTATAAGTTAACTACTGGGGATTCATTCTATGTTCCTTCAGGT AATCATTACAACATTAAAAACCTCTTGAATGTGGAAAGTAGTCTTCTTTTTACTCAAATAAAAAGGTGAAATAAATGTGTATGTATATAAAAT ACGTTGTTTATATGTACATTTTAATAACTGATTATTTTAAA
CENP-T	AGTTAGGAGGGGTGGGGTCCTGAACATGGCGGACCTCAGCTTTTCTGACGGCGATCCGACTGTACGCACGCTTTTGCGGCGCGTGCTGGAG ACGGCGGATTCGCGCACTCCGATGAGACGCCGGAGCACTCGGATCAATGCACAAAGAAGACGGTCTCCAACGCCATACTCTAACAGGCAG GCTAGCCAAACAAAGACAAGTGCTAGGAGGCAATCCCATGGAGCTAGGTCAGTTGGCAGATCGACTCGTGTTCAAGGTCGTGGGCGCCTG GAGGAACAGACACCTCGGACTCTGCTAAGGAATATCTTACTAACTGCCCCAGAATCGTCCACTGTGATGCCAGACCCAGTGGTGAAGCCAG CACAAGTACCAGAAGTGGCCCGGTCCTCCAGACGGGAGAGCAGTCGGGGAAGCCTGGAGCTACACCTTCCTGAACTGGAGCCCCCCTCAAC CCTGGCTCCGGGTCTAACAGCCCCTGGCAAAAGGAAACAGAAACTGAGATTATCCGTATTTCAGCAGGAAGTGAACCAGGGGTTGCCTCTC TCCCAAGAGCCACGTAGGAGTAGGAGTGCTGATGTTTCATCCTTGGCCAGCTCCTTCAATCTCACCTTTGTTCTACCTGGTCAGCCAGAGACA GTGGAGAGACCTGGTCTGGCTCGAAGACGTCCCATCCGCCAGCTTGTGAATGCTGGTGCTCTTTTGCAGGATCTGGAAGACAATTCCTTAGC CTCTGCTCTTCCAGGTGACAGTCACAGAACACCTGTTGCTGCTCTGCCGATGGATGTAGGCTTGGAAGACACACAGCCTTTCTCACAGTCCTT GGCTGCCTTTTCCCTTAGTGGGAAACATTCTCTGCCCAGCCCTTCTCGCCCTGGGGTTGAAGATGTTGAAAGAGTTGTGGGCCCACCAAGCT CGGGGACCAGGCTGCAAAGCCGCATGAGCAGAACTGGCTCTGCTGCCTCTCCTTTCCCCTTCCTGGAGCCACAGCCTCAGCCTGCAGAACCG AGAGAAGCAGTAGGATCCAACGAGGCTGCGGAGCCCAAGGATCAAGAAGGATCTTCAGGGTATGAGGAAACTTCTGCTAGGCCAGCAAGT GGAGAATTATCCTCCAGCACACATGACTCACTCCCAGCTGAACAGCCACCCCTGTCACCTGGAGTTGCAGTCCTATCTTCAGAGCCCCTGGA GTCTGTGACAGCCAAGTGCCCATCCAGGACCCAAACTGCAGGTCCTAGGCGCCGCCAAGACCCCCACAAGGCTGGACTGAGCCCCTATGTG AAGTTCTTCAGCTTCTGTACTAAGATGCCCGTGGAGAAGACGGCTCTTGAGATAGTAGAGAAGTGCCTAGACAAGTATTTCCAGCACCTTTG CAATGACCTGGAGGTATTTGCTGCTCATGCTGGCCGTAAGATTGTGAAGCCAGAGGACTTATTGCTGCTGATGCGACGGTAGTGANTAGAA

	ATGGGAGTCCGAGCATGGGCGGGCTGGTGCTTATCTGCCCTTGGAGTACCGGCAACAACTTATCCCNNGTGTGCATNNAGNCGGCNGATT CNNNCNCTCCGATGAGACGCCGGAGCGCTCGGATCAA
HJURP	ATGGAGTCTATGGGTCGGCAGGACCGCAGGCTGCATCAGCAGCTTAATGAAAGCAGTAGCCGCTTCCAGACGTTAATGAAGCGGCTGATA GCGAAGTACAACCAGCCCTTTGAGGACGACCCGTTGGTGGAGATGACGACGTTGACCTACGAGACGCCCCAGGGGTTGAGAGTTTGGGGA GGAAAGCTGATGAAGAAGGAAGACAAGGAATACACTCAGGTGATTGACAGATTAAATGGGCAAGCTCCAGAAGGTGATTCAGAGAGCAG CGGTGCGGATACAAGCTTAGAGGAGAACTGGCCATCTTGCAGCTCAGCGATGAGAGAAGCCTCTGGTGATCCTAGGCAGCGTCAGCCTGC AGTGCCCGGAAACACCTTGGAAACTGATTTAAGAAGGAAGTACTTGACCCAGGTGGATATATTGCTTCAAGATGAAGAGTATTTTAAGAAC GCAGAGAAGAGAGGTGGAAAAGACACAGTCATGACTTGGGTTCCTTCCGTGACCTCATCTGTCACACCAGCCTCTGGATGCCAGGATGCTA TCTCTGCAAAGAGCTCTGGTGGTCCTGGAGTATCTGCTTTGTCCTCTAGAGGCCAGGGTCCCTCATACCCTTGCCCAGCTGACATGGCCATTG TGGCTGAAGTGATGGCCTCTCCTTACTTGGGACCAGCAGCAACAGCGTAAGCAGCCAGTCCTTTGAGGTTGATGACCTTTGCAATGTGACAA TCAGCGATTTGTATGAAGGCATGATGCATTCCATGAGTAGGCTGCTGCGCTCAAAGCCATCCTGCATCATCTCCACCAAGACCTACATCAACC AGAGCTGGAAGCTCCGGAGGAGACCCTCACGCAAGCAGGGGCTGCACAAGAACAGGACAACTGCCCCCGGAGCAAGCCCTCCCAGAGGA GTGCCAGGAAGGGGCCTGCGCCCTGCTCAGAGCCCGGGAAAGAGGCCGGAATATTAAGAGACTACGGGAATTTACTGCACGTTGCTCCCC ACAAGACAGGCTTAGAACTGAAAAGCGTTTCTCTTGAAGAAGCAAACGCCAAGTCCATAAATTCAGTCCAGCTTGGAAGGAGCTCCAGATG ATGCCGCAGAAGGACTTGGACTTAAACCGTGAGCGAGAAAATAGGGTGATGACATTACAGTGGCTAATTTCACCCATCAAGGTGGTTCCCA GACCAAGGATGCTTCCAAGTCAAGTAGAGAAGTGGTATAGAGAGATTAAAATCAAATTTGACAAGCTTCATCAGGAATATTGCCTGTCTTCT GGGAAACAGCCTCACCTGACTGGCCCTACAGAGTCTTGGGTTGTCGATGTGTACAGAAGTGGTTCTAAGAGCCCAGGCAGCCGTCGGGATG TAGAAACCTGCAGGCCCAGTTCACCTTTCGGCAGAGAAAAGACAGAAAGGCCAGGTGAAGCTTTAGAAGACCTGAGGGGAAATGGAAAGT CTGTCAAAACAAAGAGCTGTCTTCTAAGGAGCTGTCCCTCTCCAGAGGGCAGCCCGTCACGGAGCCCCGGCCATTCTCAGCTGAGCTCTGGC CTTCAAGAGCACAATTCTGAACCTACTGGTAAGGCAGTGTGGCCCAGTACAGCTATTTCAGCACCATGCATAGGCTCTCCAGGCTGTGGGAA GGATAACTATGATGAGCTTAAAAAAGAATTCAACAGACTTTATCAGAAGTATTGTCTGTCACCGCAGCGGGCAAAGGTGACTTCGTGTGGC AGAGTATCGCCAATGAAAGCTGCTGCAGCTCTTCCCTGTCAATCAGAGCACTTGAAGAGACTAAATCCAGACTCCCCACAACAGAGTTCCCA GAAACGGTCAATCTCTCCTGGCTGTCACAGAAGAGTCCTGCAGGATTCAACTGCACAGACTGCTAGTACACTTGTGAGAGAGCCTTGGCTGC CAACAAAGAGGCGCAAGTTGTCATACCCGGTAGCTTGTGCTCACCAGGCCAAATCTCACAATACCTCAGGAGCATCAGGCTGGCCCTGAGT GTCATGAGTGCTGATTGGGTGTTTGTTGGACATGATGCTTATGTGTAGTCAGCAGCTGCTTTGA
Mis12	ATGTCTGTGGATCCGATGGCCTATGAGGCCCAGTTCTTTGGCTTCACACCACAGACTTGCCTGCTGAGGATCTACGTAGCATTTCAAGACCAC CTGTTTGAAGTGATGCAGGCTGTTGAACAGGTTATCCTAAAGAAGCTGGAGGACATCCCAAACTGTGAGATCACCCCTGTCCAGACTCGGA AATGCACAGAGAAGTTTCTTTGCTTCATGAAAGGGCGCTTCGATAACCTTTTTGGCAAAATGGAGCAGCTGATTTTGCAGTCGATTTTGTGTA TTCCCCCAAACATCCTGCTTCCTGAAGACAAGTGTCAGGAGATGAATCCTTTCAGTGAAGAAAAACTCCAGCTTCTCCAACAGGAAATCAAA GAGTTACAGGAGAAATACAAGGTTGAGTTGTGCACTGAGCAGGCCCTTCTTGCAGAATTAGAGGAACAGAAAACTGTTAAGGCCAAACTCA GAGAGACCTTAACTTTCTTTGATGAGCTTGAAAACATCGGCAGATATCAAGGAACTAGTAACTTTAGGGAGAGTTTGGCATCCCTGGTCCAG AGCTGCAGAAAACTTCAGAGCATTAGAGACAATGTAGAAAAAGAAAGCAGGAGACTGGAAACACAGTGA
Mis18	ATGTCGAGCGAGTCCCCCCTGCTGGAGAAGCGTCTGTCGGAAGACTCGAACCGCTACCTGCGGCTGCAGAAGTGGGCAAACATGTCGAGC GCAGACGCGTTAGGGCTAGAGAAGGAACGGCCTGAGGAGAAGGCGGCCGCGGCGGAGAACCCGCTGGTGTTCCTTTGTGCCCGCTGTCG CCGGCCACTGGGCGACTCGCTCACCTGGGTGGCCAGCCAGGAGGACACCAACTGCATCCTGTTGCGCAGTGTCTCCTGTAACGTCTCTGTG GATAAGGAACCGAAACTGTCCAAGTGCAGAGATGAAGACGGTTGCATCCTTGAGGCGCTGTACTGCACGGGCTGCTCCCTCAGCCTTGGCT ATGTGTACAGATGCACTCCTAAGAACCTGGACTACAAGCGCGACTTGTTCTGCCTCAGCGTTGAAGCCGTTGAAAGTTATACCTTAGGGTCC TCTGAGAAGCAAATCGTGTCAGAAGACAAGGAGCTTTTCAACCTTGAAAGCCGAGTTGAAATAGAGAAGTCTATAAAGCAGATGGAAGAA GTCCTGACAGCCCTGCAAAAGAAGCTACGGGAAGTTGAATCTAAACTGTCCTTGGCCAGCCGGGGCAGCTGA
Mis18bp	ATGATTGTAACACCTTTGAAGCATTCAGGAATTCATTTGTCTTCTGGAACATTGCAAAGAAGAAATATGCCTCTAGATGCAGTCTTTATTGAC AGCATCCCTTCAGGCACACTCACTCCTTTGAAAGATTTGGTAAAATATCAGAAGTCCTCTCTAAAAGTTAATGGTCATAAAAAGAATCAGTTA TTAGAAATAAGAACTTCTAACAATAAAGATCTATTTCAATCCACCATGTTATCAGAGGCTACCCTTCCAAACACCTCTCTTGATATCAGTGTTA TAAAAAAGCCCAACATGGATAGATTAAGAAATGAAATGATTTATGAATCACCAGGTAAAATCTTTCAAAGAATGAAAGCAAAAGTACAGCG TGACAAACAAGAACAACTGACAAGGAGCAGTAGCATGTTGGGATCACCACAGGGTGAACACAGCAAAGACTTTCCTCCTAACACGGATAAG AAAGCTCAGTTGCAACAAACCTATATCTGTGAGGAAAAACAAACGTCGGTCCAATCCAATGACCCTTCATTGGGAGATCCTCCAATTCTGAA CCAAGAACAAAAAAATGTTTCAGCTTCGTGCATTTCAAAGAAGGCATTAACTAGAGCCCAGTTTGGCGGACAAGTTCTTCATTCAAAAGAGA GTCCTGTTAGGATCACTGTGTCCAAAAAGAACACATTTGTTTTGGGAGGCATTGATTGTACTTATGAAAAGTTTGAAAATACTGATGTTAACA CCATTAGTTCTCTTTGTGTTCCTATTAAGAATCATAGCCAATCTATTACTTCTGATAATGATGTGACAACAGAAAGGACTGCAAAAGAGGATA TTACAGAACCAAATGAAGAGATGATATCCAGAAGGACTATTCAAGATCCCATAAAGAATACATCTAAAATTAAACGTTCAAGTCCAAGACCT AATTTAACACTATCTGGCCGGTCTCAAAGAAAATGTACAAAGCTTGAAACTGTTGTAAAAGAAGTAAAAAAATATCAGGCAGTCCAGCTACA GGAATGGATGATTAAAGTCATCAATAATAATACTGCTATATGTGTAGAAGGAAAGCTGGTAGATATGACTGATGTTTATTGGCATAGCAATG TAATTATAGAGCGGATTAAACACAATGAACTTAGGACCTTATCAGGCAACATTTATATCTTAAAAGGATTGATAGACTCGGTCTCCATGAAA GAAGCAGGATATCCCTGTTATCTCACAAGAAAATTTATGTTTGGATTTCCCCACAACTGGAAGGAACACATTGATAAATTTCTAGAACAATTA AGGGCTGAAAAAAAGAACAAGACCAGACAGGAAACAGCAAGAGTCCAAGAAAAACAAAAATCAAAAAAAAAAGATGCAGAAGATAAAGA AACTTATGTCCTGCAAAAGGCCAGCATCACGTATGACCTGAATGATAATAGCGTAGAGAGAACTGAAGTACCCACTGATCCCTTGAACTCAC TGGAACAGCCTACCTCCGGCAAAGAAAGAAGACACCCGCTTCTCAGTCAGAAGAGAGCTTATGTTTTAATAACACCACTTAGAAACAAAAA GTTGATAGAGCAAAGATGTATAGACTACAGTGTCTCTATTGAAGGAATATCGGACTTTTTCAAAGCAAAGCATCAAGAAGAAAGTGACTCA GATATACATGGAACTCCAAGTTCTACCAGTAAGTCTCAAGAGACCTTTGAACATAGAGTGGGATTTGAAGGCAATACCAAGGAGGACTGCA ATGAATGTGACATAATCACTGCCAGACATATTCAGATACCTTGCCCGAAAAGTAAACAAATGCTCACCAATGATTTTATGAAAAAGAACAAG TTGCCCTCAAAACTGCAGAAAACTGAAAGTCAAATAGGTGTATCACAGTATTGCCAGTCCTCATCACATTTGTCAAGTGAAGAGAATGAAGT AGAAATTAAAAGTAGAACCAGAGCAAGAAATACAAAAGAAAGGCTGAATCGGGAGAGAGAAAACACAAACCACATCACAAAAGATATCCT CCTCATTTCAGAAACTGAAGGTGAGAGAGCTTGTTATATCACACCAAAGAGACCTCGATCTTCTGCTAAAGAATCACATTACAAATCTGCTGT TAGCAAGGACTTTCTCATTGAAGGAAAGGCATCTGATAGAACTAGCAGACAGCTGTTAGACCATCTACCTGGTTTAACCGATGATGAAGAAT GGAGTGAGCAAGAGTTACAGAAGCTACATTGTGCTTTTACATCTCTTCCAAAGCACAAACCTGGCTTTTGGTCAGATGTGGCTATGGCAGTA GGTTCTCGAACTGCTGATGAGTGCCAGAAGAAATACACAGAAGAACCCCAAGGCCAAGGATCCCGAAAACATGTCTCTAAGAAAAAGCAA GCCAACAAGGTCCAAAATGGTGAGAAAGACAGTGCTGATGCCAAGACTATTAAGATAACTGCCAAAGTGGGAACTCTTAAAAGGAAACGA

	CAAATGAGGGATTGCCTAGAACATTTGGCAAAAGACAACCATGATGATTTTTTCACTGCAACACCCTTGCAGAAACAAAGAATTCAGTTGCC AACTTTTCAATACAGTCAAGATGATGATTTTCTGCTAGATATGGACAGGGATCCAGCATCTCCATCATCAATTATTACTTCACCATTGAGAAG CACTACTCCACAATGTCGGCATTTCAGCCCCAGCATGCTAGCCGCTATAGAAAGGAATAACTGTGATAGATATGTTTACCAAATGCAAAAAA ATGCTAAAAAATATGGCAAGAGTAATGGTGGCCTTGTCTGGGGCAACATTAGGAAAAAAACAGTTAAAACTGATCTCTCATCTCCACCACCA ACACGGAAAGCCCTGTTTAACAAAGATTTAGGTAAAAACACTGATATTTCAAAGTATTTCATTGATGATACAGAATCAGATGAAGAAGAGAA AGATTATTACTTTTCCAATTCTGACTGA

Supplementary Table 2. *M. macedonicus* constructed centromeric protein coding gene transcripts

Name	Sequence
CENP-A	ATGGGCCCCCGTCGCAAACCGGAGACCCCAAGGAGGAGACCCTCCAGCCCGGCGCCTGGACCCTCGCGACAGAGCTCCAGTGTAGGCT CTCAGACACTGCGCAGAAGACAGAAATTCATGTGGCTTAAGGAAATCAAGACCCTGCAGAAGAGCACAGACCTCTTGTTCAGGAAGAAG CCTTTCAGCATGGTTGTTAGAGAAATATGTGGGAAGTTCAGCCGTGGTGTGGATTTATGGTGGCAAGCCCATGCCTTGTTGGCCCTTCAG GAGGCAGCAGAAGCTTTCCTCGTCCACCTCTTTGAGGACGCCTACCTCCTCTCCTTACATGCTGGTCGGGTCACGCTTTTCCCCAAAGACA TTCAGTTGACCAGGAGAATCCGAGGCTTCGAGGGCGGACTCCCCTAA
CENP-B1 Head	ACAGACGACTTCACTGCTTCCAACGGCTGGCTGGATCGCTTCCGCAGGCGCCACGGTGTAGTGGCCTGCAGCGGCGTGACCCGTTCTCG GGCGCGAAGCTCTGCCCCTCGGGCCCCAGCGGCACCTGCCGGCCCAGCCACGGTGCCCTCCGAGGGCAGCGGTGGAAGCACACCGGGC TGGCGCACGAGGGAGGAGCAGCCGCCGTCGGTAGCTGAGGGCTACGCCTCGCAGGACGTGTTCCGCGCCACCGAGACCAGCCTGTGGT ACGACTTTCTGTCGGACCAGGCCTCAGGGCTCTGGGGAGGTGATGGACCGGCGCGCCAGGCCACCCAGCGCCTTAGCGTTTTGCTGTGC GCCAACGCCGATGGCAGCGAGAAGCTTCCCCCACTGGTCGCAGGCAAATCTGCCAAGCCCCGTGCAGGGCAAGGTGGTCTGCCCTGCG ACTACACTGCCAACTCGAAGGGTGGAGTCACCACCCAGGCTCTGGCTAAGTACTTGAAAGCTCTGGACACTCGAATGGCAGCAGAATCT CGTCGGGTCCTTCTGCTGGCCGGCCGTCTGGCTGCCCAGTCCTTGGACACCTCGGGCCTGCGGCACGTGCAGCTGGCCTTCTTCCCTCCT GGCACCGTGCATCCTTTGGAGCGAGGAGTGGTCCAACAGGTGAAGGGCCACTACCGCCAGGCAATGTTGCTCAAGGCCATGGCAGCAC TCGAGGGTCAGGATCCCTCAGGCTTGCAACTGGGCCTTGTGGAGGCCCTGCACTTTGTGGCTGCAGCCTGGCAGGCAGTGGAGCCCTCG GACATAGCAACTTGCTTTCGTGAGGCGGGTTTTGGAGGTGGTCTTAATGCCACTATCACCACTTCCTTCAAGAGTGAGGGAGAGGAAGA GGAGGAAGAGGAGGAGGAAGAGGAGGAGGAGGAAGAGGAGGGTGAAGGGGATGAGGAGGAGGANGAGGAAGAGGAAGGGGAG GAGGAAGAAGGAGGGGAAGGAGAGGAGGTAGGAGAGGAAGAGGAAGAGGTAGAAGAGGAGGGTGAGGTTGATGATAGTGATGAA GAGGAGGAAGAAAGTTCCTCTGAGGGTTTAGAGGCTGAAGACTGGGCCCAGGGAGTAGTGGAGCCAGTGGTGGTTTTGGGGGTTATA GTGTCCAGGAAGAGGCCCAGTGCCCCACTCTCCATTTCCTGGAAGGAGGAGAGGACTCTGACTCAGACAGTGATGAAGAGGAGGACGA TGAAGAGGAGGATGAGGATGATGATGATGAGGATGATGATGAGGATGGTGATGAGGTCCCTGTGCCCAGCTTTGGGGAGGCCATGGC TTACTTTGCCATGGTCAAGAGGTACCTGACCTCCTTCCCCATCGATGACCGAGTGCAGAGTCACATCCTTCATTTGGAACACGATCTGGTC CACGTGACTAGGAAGAACCATGCCAGGCAGGCGGGAGTTCGGGGTCTTGGACATCAAAGCTGA
CENP-B2 Tail	ATGGACGACTTCACTGCTTCCAACGGCTGGCTGGATCGCTTCCGCAGGCGCCACGGTGTAGTGGCCTGCAGCGGCGTGACCCGTTCTCG GGCGCGAAGCTCTGCCCCTCGGGCCCCAGCGGCACCTGCCGGCCCAGCCACGGTGCCCTCCGAGGGCAGCGGTGGAAGCACACCGGGC TGGCGCACGAGGGAGGAGCAGCCGCCGGCGAGGTAGCTGAGGGCTACGCCCTGCGGCAGGGACGGGCGCAGCGCCACCGAGACCAG CCTGTGGTACGACTTTCTGTCGGACCAGGCCTCAGGGCTCTGGGGAGGTGATGGACCGGCGCGCCAGGCCACCCAGCGCCTTAGCGTTT TGCTGTGCGCCAACGCCGATGGCAGCGAGAAGCTTCCCCCACTGGTCGCAGGCAAATCTGCCAAGCCCCGTGCAGGGCAAGGTGGTCT GCCCTGCGACTACACTGCCAACTCGAAGGGTGGAGTCACCACCCAGGCTCTGGCTAAGTACTTGAAAGCTCTGGACACTCGAATGGCAG CAGAATCTCGTCGGGTCCTTCTGCTGGCCGGCCGTCTGGCTGCCCAGTCCTTGGACACCTCGGGCCTGCGGCACGTGCAGCTGGCCTTCT TCCCTCCTGGCACCGTGCATCCTTTGGAGCGAGGAGTGGTCCAACAGGTGAAGGGCCACTACCGCCAGGCAATGTTGCTCAAGGCCATG GCAGCACTCGAGGGTCAGGATCCCTCAGGCTTGCAACTGGGCCTTGTGGAGGCCCTGCACTTTGTGGCTGCAGCCTGGCAGGCAGTGGA GCCCTCGGACATAGCAACTTGCTTTCGTGAGGCGGGTTTTGGAGGTGGTCTTAATGCCACTATCACCACTTCCTTCAAGAGTGAGGGAGA GGAAGAGGAGGAAGAGGAGGAGGAAGAGGAGGAGGAGGAAGAGGAGGGGTAGAAGGGGAAGANGAGGAGGAGGAAGAGGAAG ANGAAGGGGAGGAGGAAGAAGGAGGGGAAGGAGAGGAGGTAGGAGAGGAAGAGGAAGAGGTAGAAGAGGAGGGTGAGGTTGAT GATAGTGATGAAGAGGAGGAAGAAAGTTCCTCTGAGGGTTTAGAGGCTGAAGACTGGGCCCAGGGAGTAGTGGAAGCCAGTGGTGGT TTTGGGGGTTATAGTGTCCAGGAAGAGGCCCAGTGCCCCACTCTCCATTTCCTGGAAGGAGGAGAGGACTCTGACTCAGACAGTGATGA AGAGGAGGACGATGAAGAGGAGGATGAAGGATGATGATGATGAGGATGATGATGAGGATGGTGATGAGGTCCCTGTGCCCAGCTTTG GGGAGGCCATGGCTTACTTTGCCATGGTCAAGAGGTACCTGACCTCCTTCCCCATCGATGACCGAGTGCAGAGTCACATCCTTCATTTGG AACACGATCTGGTCCACGTGACTAGGAAGAACCATGCCAGGCAGGCGGGAGTTCGGGGTCTTGGACATCAAAGCTGA
CENP-C	AAGGCCAGAACATGCTGGAAATCTTACAAGATTGTTTTGAAGATCAGAGTAAGGCTTCATTTCTTGATGATTTTACAGAGTCATTGACTTC TTCAACACAAAAAAAGAAAGCTAATTATAGTCAATCATCAAGCAAAAAGTGCCCAGAGTCACATTCCAAACCAGTTCCAGTGTCTTCAAG GACGGGAGAAGCTTCTCTGCAAGCGTCTGCAGAGCCGAGTGAAGCCGCTGGTGGATCAGTTCAGGCTAATGAAGTTCACCATGGTGTGT CTGGTGAATTGGACTTATGTGTTCGCTCCCCTGTGGTTCTCCTGGATGCAAACGTCAACACGTTACAGAAAGCCGCTTCACCTGCTGGCCA AAAGAGAGTGGCTTCTGTCTCAAGGAGTCCAGTAGATAGGCAGGCTTCAAATAAAAACATTTCTTTTAAAACCAGGAAAAGGTTAAACTT TGAAGATAAAGTTACTTCGAGCACAGCAGAAACAGAGAACAGTGTGTTACAAGTAGAGGATAACTTATCCAAGGGACAAGAAGGGACA TCATCAGAAATCACTCAGAAAAGAGACGATCTGAGTTCTGATGTCCAATCACGTTCTAAGAAGAATTTTTCGGAGTTGTTTTTAGAGACT GTCAAGAGGAAAAGTAAATCCAGTTCTGTGGTCAGGCACACAGCAGCAGTTCCATCTTCTCCGCCTCCTCCAAGTGACATGAAATTATTA GAAGATGAATTTATTATTGATCGGTCAGATAGAAGTTTTTCAAGTCGACTTTGGGTGATGATACCAAGCAAAGACAGACATCTGAGCGCC CACAAGCCATCTCCTGAAAACACTGCACTGCTTCAAGGTAAGAAGTCAAGAGAGAAATCTCATAGTTTATCTGCAATGACGTTTGCAAGA AACACACAGTCTGATAAGGCTCACCCAATAGAGGAAGCTCAGCTCTCTGTTGAAGAAAACCTGGGAACTACTTGTACAGATGAGTTGGA AAATGATTGTCGATCTCCTGAAAATAAAATGCAATCTGAGACTGCCAAAACCCCACCCGCATGGGAAAGGACTACGAAACAGAACCAGA GGAGAGTGTCTAAGCCTAAAGCAGCTGAAGAGCTCAGGAAGGGACAGAGTAGCTGTGAAAATAGCAATGTGTCAAACACTGGCCAAGA CAAGCTACAGATCAACTCGAAAAGAAATATGAAGGACTGTGAAGAGGTGAGGAATGAGCCTACTCCCAAGAAACAGAAGCCAGCTCTT GAAAATAAGAAAAAAAACAATAGCACCCAAACCAATAAGGAAAAATCTGGAAAGAAATTTTTTTCTGGTGGGTCCAAGAACAAATTTGT TCCCAAAAAAGTAACTTTGACTTCCAGGAGAAGTTGTAGAATTTCTCAGCGTCCATCTGACTGGTGGAGGGTAAAATCAGATGAAAGTTC

	TGTTGATAGAAATCCTTCAAAAGAAAATAACTCACCAGTGGTATATCCCAATAAAAAGAAGCAGACTAAGAGAAATCATGTATCTAAGAG GGCTGGGAAGAAACCTGGTCCATCAAAAAGACAGAAGACAGAGATGAGCCCAAGAGTACAGAAGTCTTTAAATGTAAAGGATTCTGGA GGAACTGTTGGTGGCCATGATGACACTTCTCGTTCCCAGAGGAAGCCATTGAAAAACATTGAGGCAGACCCAACTCAGAAGAGCCTTGC TATTTCTCGACCTAAAAGAGGCTGTAAGTATCGAAACAATGTTATGACATCACCGAATGTTCATTTAAAGTCTCATACTGAGGAATATACA AGCAAGACACAAATGGAGTCTGCTTCAAATTCAGAAGTGTCTAAACGTTCAGTGTGGGAAGAAAGTGGACCTTCCAGGTTCAAGAATTA TGAAATGCCAGGGAGCAGTAATTCTGAGATGGGTGATGAACAGGACCAGAAGAGTTTGCATTTTACAACAAGAAGTTTTAATATGGTAC CAGATAAAAAATTACATCACAAATTAGTACTGCCCTCCAACTCACCAAATGTTCGTCGATCTAATAGAATACGTTTGAAACCTCTGGAGTA TTGGCGAGGGGAACGGGTAGATTATCAAGAAAGTTCGTCGGGACAACTTGTGCTTGAAATAATATCCCCAGCTTCAGTACCCACGAAAA TAAAGGCACAAAGAAACCTCGGCAAAGTCAACAAAAAAGTTACCAAGAAGCAGACCCATCTCAATAGCCATGAAAAGGCTAAGATGGA GCTGCCTCTGGATATGCGTCTGGGAGATCCCTTTCAGGCAACATTGGCGAAGGACCCAGAAACCGCAGAGCTTGTCCCCATGGATCTTAT AAGTCCACAAGATACATACCGCTTTTTTGTTGAGCAACATGGTTTGAAAGTTTTTAAAACATTGGATACAATATATTTTTCTACTGGAAAAT TGGTTTTAGGACCCTATGAAGAGAAAGGAAAACAGCATGTTGGCCAAGATATATTGGTTTTTTATGTTAACTTTGGTGACCTTTTATGTAC ATTACATGAAACACCTTATAAGTTAACTACTGGGGATTCATTCTATGTTCCTTCAGGTAATCATTACAACATTAAAAATCTCTTGAATGTGG AAAGTAGTCTTCTTTTTACTCAAATAAAAAGGTGAAATAAATGTGTATGTATATAAAATACGTTGTTTATATGTACATTTTAATAACTGATT ATTTTAAA
CENP-T	GGGACGCCTCCGTCCTCAATCTTACTCTAATCCTCCTGGAACCCTGGCCGGCTTGGGTTTCCCTCTCAGGCCTCCGTTTGACCTAGTGGGC CGGAACCATGGCCTCGTTCCATCTGGATCATCTAAAGAACTACCACAGAAGATATTGCCGGCCTTCCAGAGCACCCAACATTCATACAAA AAAAGGCCAGAACATGCTGGAAATCTTACAAGATTGTTTTGAAGATCAGAGTAAGGCTTCATTTCTTGATGATTTTACAGAGTCATTGAC TTCTTCAACACAAAAAAAGAAAGCTAATTATAGTCAATCATCAAGCAAAAAGTGCCCAGAGTCACATTCCAAACCAGTTCCAGTGTCTTCA AGGACGGGAGAAGCTTCTCTGCAAGCGTCTGCAGAGCCGAGTGAAGCCGCTGGTGGATCAGTTCAGGCTAATGAAGTTCACCATGGTG TGTCTGGTGAATTGGACTTATGTGTTCGCTCCCCTGTGGTTCTCCTGGATGCAAACGTCAACACGTTACAGAAAGCCGCTTCACCTGCTGG CCAAAAGAGAGTGGCTTCTGTCTCAAGGAGTCCAGTAGATAGGCAGGCTTCAAATAAAAACATTTCTTTTAAAACCAGGAAAAGGTTAA ACTTTGAAGATAAAGTTACTTCGAGCACAGCAGAAACAGAGAACAGTGTGTTACAAGTAGAGGATAACTTATCCAAGGGACAAGAAGG GACATCATCAGAAATCACTCAGAAAAGAGACGATCTGAGTTCTGATGTCCAATCACGTTCTAAGAAGAATTTTTCGGAGTTGTTTTTAGA GACTGTCAAGAGGAAAAGTAAATCCAGTTCTGTGGTCAGGCACACAGCAGCAGTTCCATCTTCTCCGCCTCCTCCAAGTGACATGAAATT ATTAGAAGATGAATTTATTATTGATCGGTCAGATAGAAGTTTTTCAAGTCGACTTTGGGTGATGATACCAAGCAAAGACAGACATCTGAG CGCCCACAAGCCATCTCCTGAAAACACTGCACTGCTTCAAGGTAAGAAGTCAAGAGAGAAATCTCATAGTTTATCTGCAATGACGTTTGC AAGAAACACACAGTCTGATAAGGCTCACCCAATAGAGGAAGCTCAGCTCTCTGTTGAAGAAAACCTGGGAACTACTTGTACAGATGAGT TGGAAAATGATTGTCGATCTCCTGAAAATAAAATGCAATCTGAGACTGCCAAAACCCCACCCGCATGGGAAAGGACTACGAAACAGAAC CAGAGGAGAGTGTCTAAGCCTAAAGCAGCTGAAGAGCTCAGGAAGGGACAGAGTAGCTGTGAAAATAGCAATGTGTCAAACACTGGCC AAGACAAGCTACAGATCAACTCGAAAAGAAATATGAAGGACTGTGAAGAGGTGAGGAATGAGCCTACTCCCAAGAAACAGAAGCCAGC TCTTGAAAATAAGAAAAAAAACAATAGCACCCAAACCAATAAGGAAAAATCTGGAAAGAAATTTTTTTCTGGTGGGTCCAAGAACAAATT TGTTCCCAAAAAAGTAACTTTGACTTCCAGGAGAAGTTGTAGAATTTCTCAGCGTCCATCTGACTGGTGGAGGGTAAAATCAGATGAAAG TTCTGTTGATAGAAATCCTTCAAAAGAAAATAACTCACCAGTGGTATATCCCAATAAAAAGAAGCAGACTAAGAGAAATCATGTATCTAA GAGGGCTGGGAAGAAACCTGGTCCATCAAAAAGACAGAAGACAGAGATGAGCCCAAGAGTACAGAAGTCTTTAAATGTAAAGGATTCT GGAGGAACTGTTGGTGGCCATGATGACACTTCTCGTTCCCAGAGGAAGCCATTGAAAAACATTGAGGCAGACCCAACTCAGAAGAGCCT TGCTATTTCTCGACCTAAAAGAGGCTGTAAGTATCGAAACAATGTTATGACATCACCGAATGTTCATTTAAAGTCTCATACTGAGGAATAT ACAAGCAAGACACAAATGGAGTCTGCTTCAAATTCAGAAGTGTCTAAACGTTCAGTGTGGGAAGAAAGTGGACCTTCCAGGTTCAAGAA TTATGAAATGCCAGGGAGCAGTAATTCTGAGATGGGTGATGAACAGGACCAGAAGAGTTTGCATTTTACAACAAGAAGTTTTAATATGG TACCAGATAAAAAATTACATCACAAATTAGTACTGCCCTCCAACTCACCAAATGTTCGTCGATCTAATAGAATACGTTTGAAACCTCTGGA GTATTGGCGAGGGGAACGGGTAGATTATCAAGAAAGTTCGTCGGGACAACTTGTGCTTGAAATAATATCCCCAGCTTCAGTACCCACGA AAATAAAGGCACAAAGAAACCTCGGCAAAGTCAACAAAAAAGTTACCAAGAAGCAGACCCATCTCAATAGCCATGAAAAGGCTAAGAT GGAGCTGCCTCTGGATATGCGTCTGGGAGATCCCTTTCAGGCAACATTGGCGAAGGACCCAGAAACCGCAGAGCTTGTCCCCATGGATC TTATAAGTCCACAAGATACATACCGCTTTTTTGTTGAGCAACATGGTTTGAAAGTTTTTAAAACATTGGATACAATATATTTTTCTACTGGA AAATTGGTTTTAGGACCCTATGAAGAGAAAGGAAAACAGCATGTTGGCCAAGATATATTGGTTTTTTATGTTAACTTTGGTGACCTTTTAT GTACATTACATGAAACACCTTATAAGTTAACTACTGGGGATTCATTCTATGTTCCTTCAGGTAATCATTACAACATTAAAAATCTCTTGAAT GTGGAAAGTAGTCTTCTTTTTACTCAAATAAAAAGGTGAAATAAATGTGTATGTATATAAAATACGTTGTTTATATGTACATTTTAATAACT GATTATTTTAAA
HJURP	ATGGAGTCTATGGGTCGGCAGGACCGCAGGCTGCATCAGCAGCTTAATGAAAGCAGTAGCCGCTTCCAGACGTTAATGAAGCGGCTGA TAGCGAAGTACAACCAGCCCTTTGAGGACGACCCGTTGGTGGAGATGACGACGTTGACCTACGAGACGCCCCAGGGGTTGAGAGTTTG GGGAGGAAAGCTGATGAAGAAGGAAGACAAGGAATACACTCAGGTGATTGACAGATTAAATGGGCAAGCTCCAGAAGGCGATTCAGA GAGCAGCGGTGCGGATACAAGCTTAGAGGAGAACTGGCCATCTTGCAGCTCAGCGATGAGAGAAGCCTCTGGTGATCCTAGGCAGCGT CAGCCTGCAGTGCCCGGAAACACCTTGGAAACTGATTTAAGAAGGAAGTACTTGACCCAGGTGGATATATTGCTTCAAGATGAAGAGTA TTTTAAGAACGCAGAGAAGAGAGGTGGAAAAGACACAGTCATGACTTGGGTTCCTTCCGTGACCTCGTCTGTCACACCAGCCTCTGGAT GCCAGGATGCTATCTCTGCAAAGAGCTCTGGTGGTCCTGGAGTATCTGCTTTGTCCTCTAGAGGCCAGGGTCCCTCATACCCTTGCCCAG CTGACATGGCCATTGTGGCTAGAAGTGATGGCCTCTCCTTACTTGGGACCAGCAGCAACAGCGTAAGCAGCCAGTCCTTTGAGGTTGAT GACCTTTGCAATGTGACAATCAGCAATTTGTATGAAGGCATGATGCATTCCATGAGTAGGCTGCTGCGCTCAAAGCTATCCTGCATCATCT CCACCAAGACCTACATCAACCAGAGCTGGAAGCTCCGGAGGAGACCCTCACGCAAGCAGGGGCTGCACAAGAACAGGACACACTGCCC CCGGAGCAAGCCCTCCCAGAGGAGTGCCAGGAAGGGGCCTGCGCCCTGCTCAGAGCCCAGGAAAGGGGCCGGAATATTAAGAGACTA CGGGAATTTACTGCACGTTGCTCCCCACAAGACAGGCTTAGAACTGAAAAGCGTTTCTCTTGAAGGAAGCAAACGCCAAGTCCATAAATT CAGTCCAGCTTGGAAGGAGCTCCAGATGATGCCGCAGAAGGACTTGGACTTAAACCGTGAGCGAGAAAATAGGGGGATGACATTACAG TGGCTAATTTCACCCGTCAAGGTGGTTCCCAGACCAAGGATGCTTCCAAGTCAAGTAGAGAAGTGGTATAGAGAGATTAAAATCAAATTT GACAAGCTTCATCAGGAATATTGCCTGTCTTCTGGGAAACAGCCTCACCTGACTGGCCCCACAGAGTCTTGGGCTGTCGATGTGTACAGA AGTGGTTCTAAGAGCCCAGGCAGCCGTCGGGATGTAGAAACCTGCAGGCCCAGTTCACCTTTCGGCAGAGAAAAGACAGAAAGGCCAG GTGAAGCTTTAGAAGACCTGAGGGGAAATGGAAAGTCTGTCAAAACAAAGAGCTGTCTTCTAAGGAGCTGTCCCTCTCCAGAGGGCAGC CCGTCACGGAGCCCCGGCCATTCTCAGCTGAGCTCTGGCCTTCAAGAGCACAATTCTGAACCTACTGGTAAGGCAGTGTGGCCCAGTACA

	GCTATTTCAGCACCATGCATAGGCTCTCCAGGCTGTGGGAAGGATAACTATGATGAGCTTAAAAAAGAATTCAACAGACTTTATCAGAAG TATTGTCTGTCACCGCAGCGGGCGAAGGTGACTTCGTGTGGCAGAGTATCGCCAATGAAAGCTGCTGCAGCTCTTCCCTGTCAATCAGAA CACTTGAAGAGACTAAATCCAGACTCTCCACAACAGAGTTCCCAGAAACGGTCAATCTCTCCTGGCTGTCACAGAAGTGTCCTGCAGGAT TCAACTCCACAGACTGCTAGTACACTTGTGAGAGAGCCTTGGCTGCCAACAAAGAGGCGCAAGTTGTCATACCCGGTAGCTTGTGCTCAC CAGGCCAAATCTCACAATACCTCAGGAGCATCAGGCTGGCCCTGAGTGTCATGAGTGCTGATTGGGTGTTTGTTGGACATGATGCTTATG TGTAGTCAGCAGCTGCTTTGA
Mis12	ATGTCTGTGGATCCGATGGCCTATGAGGCCCAGTTCTTTGGCTTCACACCACAGACTTGCCTGCTGAGGATCTACGTAGCATTTCAAGACC ACCTGTTTGAAGTGATGCAGGCTGTTGAACAGGTTATCCTAAAGAAGCTGGAGGACATCCCAAACTGTGAGATCACCCCTGTCCAGACTC GGAAATGCACAGAGAAGTTTCTTTGCTTCATGAAAGGACGCTTCGATAACCTTTTCGGCAAAATGGAGCAGCTGATTTTGCAGTCGATTT TGTGTATTCCCCCAAACATCCTGCTTCCTGAAGACAAGTGTCAGGAGACGAATCCTTTCAGTGAAGAAAAACTCCAGCTTCTCCAACAGG AAATCAGAGAGTTACAGGAGAAATACAAGGTTGAGTTGTGCACTGAGCAGGCCCTTCTTGCAGAATTAGAGGAACAGAAAACTGTTAA GGCCAAACTCAGAGAGACCTTAACTTTCTTTGATGAGCTTGAAAACATCGGCAGATATCAAGGAACTAGTAACTTTAGGGAGAGTTTGGC ATCCCTGGTCCAGAGCTGCAGAAAACTTCAGAGCATTAGAGACAATGTAGAAAAAGAAAGCAGGAGACTGGAAACACAGTGA
Mis18a	ATGTCGAGCGAGTCCCCCCTGCTGGAGAAGCGTCTGTCGGAAGACTCGAGCCGCTACCTGCGGCTGCAGAAGTGGGCAAACATGTCGA GCGCAGACGCGTTAGGGCTAGAGAAGGAACGGCCTGAGGAGAAGGCGGCCGCGGCGGAGAACCCGCTGGTGTTCCTTTGTGCCCGCT GTCGCCGGCCACTGGGCGACTCGCTCACCTGGGTGGCCAGCCAGGAGGACACCAACTGCATCCTGTTGCGCAGTGTCTCCTGTAACGTC TCTGTGGATAAGGAACCGAAACTGTCCAAGTGCAGAGATGAAGACGGTTGCATCCTTGAGGCACTGTACTGCACGGGCTGCTCCCTCAG CCTTGGTTATGTGTACAGATGCACTCCTAAGAACCTGGACTACAAGCGCGACTTGTTCTGCCTCAGCGTTGAAGCCGTTGAAAGTTATAC CTTAGGGTCCTCTGAGAAGCAAATCGTGTCAGAAGACAAGGAGCTTTTCAACCTTGAAAGCCGAGTTGAAATAGAGAAGTCTATAAAGC AGATGGAAGAAGTCCTGACAGCCCTGCAAAAGAAGCTACGGGAAGTTGAATCTACACTGTCCTTGGCCAGCCGGGGCAGCTGA
Mis18bp1	ATGATTGTAACACCTTTGAAGCATTCAGGAATTCATTTGTCTTCTGGAACATTGCAAAGAAGAAATATGCCTCTAGATGCAGTCTTTATTG ACAGCATCCCATCAGAGGCTACCCTTCCAAACACCTCTCTTGATATCAGTGTTATAAAGCCCAACATGGATAGATTAAGAAATGATATGAT TTATGAATCACCAGGTTAAATCTTTCTAAGAATGAAAGCAAAAGTACAGCGTGACTAAGAAGAGAAACTTTCCTCCAATTCTGANCCAAG AACAAAAAAAATGTTTCAGCTTCGTGCANTTCAAAGAAGGCATTAACTAGAGCCCAGTTTGGCGGACAAGTTCTTCATTCAAAAGAGAGTC CTGTTAGGATCACTGAGTCCAAAAAGAACACATTTGTTTTGGGAGGCATTGATTGTACTTATGAAAAGTTTCAAAATACTGATGTTAACAC CATTAGTTCTCTTTGTGTTCCTATTAAGAATCATAGCCAATCTATTACTTCTGATAATGATGTGACAACAGAAAGGACTGCAAAAGAGGAT ATTACAGAACCAAATGAAGAGATGATATCCAGAAGGACTATTCAAGATCCCATAAAGAATACATCTAAAATTAAACGTTCAAGTCCAAGA CCTAATTTAACACTATCTGGCCGGTCTCAAAGAAAATGTACAAAGCTTGAAACTGTTGTAAAAGAAGTAAAAGAATATCAGGCAGTCCAG CTACAGGAATGGATGATTAAAGTCATCAATAATAATACTGCTATATGTGTAGAAGGAAAGCTGGTAGATATGACTGATGTTTATTGGCAT AGCAATGTAATTATAGAGCGGATTAAACACAATGAACTTAGNACCTTATCAGGCANCATTTATATCTTAAAAGGATTGATAGACTCGGTC TCCATGAAAGAAGCAGGATATCCCTGTTATCTCACAAGAAAATTTATGTTTGGATTTCCCCACAACTGGAAGGAACACATTGATAAATTTC TAGAACAATTAAGGGCTGAAAAAAAGAACAAGACCAGACAGGAAACAGCAAGAGTCCAAGAAAAACAAAAATCAAAAAAAAAAAGAT GCAGAAGATAAAGAAACGTATGTCCTGCAAAAGGCCAGCATCACGTATGACCTGAATGATAATAGCGTAGAGAGAACTGAAGTACCCA CTGATCCCTTGAACTCACTGGAACAGCCTACCTCCGGCAAAGAAAGAAGACACCCGCTTCTCAGTCAGAAGAGAGCTTATGTTTTAATAA CACCACTTAGAAACAAAAAGTTGATAGAGCAAAGATGTATAGACTACAGTGTCTCTATTGAAGGAATATCGGACTTTTTCAAAGCAAAGC ATCAAGAAGAAAGTGACTCAGATATACATGGAACTCCAAGTTCTACCAGTAAGTCTCAAGAGACCTTTGATCATAGAGTGGGATTTGAA GGCAATACCAAGGAGGACTGCAATGAATGTGACATAATCACTGCCAGACATATTCAGATACCTTGCCCGAAAAGTAAACAAATGCTCAC CAGTGATTTTATGAAAAAGAACAAGTTGCCCTCAAAACTGCAGAAAACTGAAAGTCAAATAGGTGTATCACAGTATTGCCGGTCCTCATC ACATTTGTCAAGTGAAGAGAATGAAGTAGAAATTAAAAGTAGAACCAGAGCAAGAAATACAAAAGAAAGGCTGAATCGGGAGAGAGA AAACACAAACCACATCACAAAAGATATCCTCCTCATTTCAGAAACTGAAGGTGAGAGAGCTTGTTATATCACACCAAAGAGACCTCGATC TTCTGCTAAAGAATCACATTACAAATCTGCTGTTAGCAAGGACTTTCTCATNGAAGNAAAGGCATCTGATAGAACTAGCAGACAGCTGTT AGACCATCTACCTGGTTTAACCGATGATGAAGAATGGAGTGAGCAAGAGTTACAGAAGCTACATTGTGCTTTTACATCTCTTCCAAAGCA CAAACCTGGCTTTTGGTCAGATGTGGCTATGGCAGTAGGTTCTCGAACTGCTGATGAGTGCCAGAAGAAATACACAGAAGAACCCCAAG GCCAAGGATCCCGAAAACATGTCTCTAAGAAAAAGCAAGCCAACAAGGTCCAAAATGGTGAGAAAGACAGTGCTGATGCCAAGACTAT TAAGATAACTGCCAAAGTGGGAACTCTTAAAAGGAAACGACAAATGAGGGATTGCCTAGAACATTTGGCAAAAGACAACCATGATGATT TTTTCACTGCAACACCCTTGCAGAAACAAAGAATTCAGTTGCCAACTTTTCAATACAGNCAAGATGATGATTTTCTGCTAGATATGGACAG GGATCCAGCATCTCCATCATCAATTATTACTTCACCATTGAGAAGCACTACTNNNNANTGTCGGCATTTCAGCCCNAGCATGNTAGNCGC TATATAAAGGAATAACTGTGATAGATATGTTTACCAAATGCAAAAAAATGCTAAAAAATATGGCAAGAGTAATGGTGGCCTTGTCTGGG GCAACATTAGGAAAAAAACAGTTAAAACTGATCTCTCATCTCCACCACCAACACGGAAAGCCCTGTTTAACAAAGATTTAGGTAAAAACN CTGATATTTCAAAGTANTTCATTGATGATACAGAATCAGATGAAGAAGAGAAAGATTATTACTTTTCCAATTCTGACTGA

Supplementary Table 3. HJURP Oligopaint library sequences

Library name	Sequence
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTGGCCAGATGATGGTCAACTGTTACAGCTGACTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGGAACCATTGTAAGGGAAGGACAGGAGAAAAGCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGATAGTTGGTCACTTGAGTTCTGTTACCACCCACGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACTAGCTATTGGCAACTAGACACCACCTCAAGTAGCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGCCTCGTGGGTGCTGCATGGCTGCAAGAACTTTCATTGCACCCTTGCACGATAC

HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAAAGCATCAAGACCTTGGGATTTCTTCTTGTGGGTGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAGCTCAGGCAGTTTCAGCTGGCTGTCTCTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAAACACAAAATGTGGAGCATACTCTAGACGGCATGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCAGAGTCAGAGAAGACAAACAGGTCCGCAGGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAACCATATACCTTTATTTGCAAAACCACCCCAGGCACTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTTCCACGGTACCAGCTGCCCTGGAGCATGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGAGGTGCATCTGATGCCAGCAAGGTGATCGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGACAGAGACCCCTTCTAAAAGCAGAACTAGAGGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTTTCATCCTGACTGCCCAGTGTAAACCCCGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTCTCCTGTTGTCAAAGAACTCAAGGCACACTACGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGCAGTTTGCCCTGGGTATGTCAGCCCAATTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTAAGAGAGACTGCCCACATCACCAAAGGGTCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAAAGCCTTACTGGTCTAGCACAAAGGAGGAAGAACTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTCCTTCATGTTGTCACTGAATGTAGAGACCTGGTCCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTAATGTGGCTCAAGATGGAACTAGAGAAGGCATCTGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTAAAGGCCCTAAGGATAGGCTAGTAACTTCCAGGGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGTGGTGAGGAACTCCTGGGTACCAGGGAATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGTAGTTGGGCAAGCGAGGTCTCCGGGCTGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCAGGAAAAGCAAGGTTTTCTCCAGGTAGAGTGCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCCAGATGGGTGGGCAAGGGACACAGCTGTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATTCTTGGGACAGCTGGGGACACAGGCTGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCACTCTCACAGTCACTTCCATAGTGATTTCTGGAGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGGCCACCAAGGCACACCACGATAGCAGTTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGGGCTCACGAGGTTGAGGCAGCTAGACAATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACAGCTCAAGGCCAAGAGCAAAGGGAGATAGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCACACCAGGCAGCCAGAGACCCAGTTCAACTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCCCACAACCAGCGCTCTTAAGACTAAGCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGAGAGGTGTGAACAGCGTTCTGTCATGACTCAACTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCACCTACTCAGAAGCAGAAGTGAGGAAGGTGAAGATGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTCATGTTAAGAGGAAAACTGCCACTGAGAATCCCCTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATAAGATAAAGAGACACAGTAATGGCCAAGGCTGGCTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGAGAAGCCCAAGGCACCAACATCCCACAATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCAGGGAGCTGCAATGGAAGACTGCTCAGTACTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACAGGATGCCCTGGGGACCGTCCTCCAAAATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGAGAAACAAAAGTCTCTGCTGAGGTCTTCTTCCTCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAGCAGCTGCTGACTACACATAAGCATCATGTCCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCGTGAAACTTGGCCTGGTGAGCACAAGCTATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGCCAAGAGTCTCTCACAAGTGTACTAGCAGTCTGTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGCAGTTGAATCCTGCAGGACTCTTCTGTGACAGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTAGTCTCTTCAAGTGCTCTGATTGACAGGGAAGAGCTTTTCATTGCACCCTTGCACGATAC

HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGAAGTCACCTTCGCCCGCTGCGGTGACAGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTAAGCTCATCATAGTTATCCTTCCCACAGCCTGGAGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCACACTGCCTTACCAGTAGGTTCAGAATTGTGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTCTTGAAGGCCAGAGCTCAGCTGAGAATGGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGGCTGCCCTCTGGAGAGGGACAGCTCCTTATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGCTCTTTGTTTTGACAGACTTTCCATTTCCCCTCAGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCGAAAGGTGAACTGGGCCTGCAGGTTTCTATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGGGTCAGTCAGGCGAGGCTGTTTCCCAGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACCACTTCTCTACTTGACTTGGAAGCATCCTTGGTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCTGGGAACCACCTTGACGGGTGAAATTAGCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACTGTAATGTCATCACCCTATTTTCTCGCTCACGGTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGAGAAACGCTTTTCAGTTCTAAGCCTGTCTTGTGGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATTCCCGTAGTCTCTTAATATTCCGGCCTCTTTCCCGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCCTTCCTGGCACTCCTCTGGGAGGGCTTGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCTGCTTGCGTGAGGGTCTCCTCCGGAGCTTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTGATGCAGGATGGCTTTGAGCGCAGCAGCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACTCATGGAATGCATCATGCCTTCATACAAATCGCTGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGCTGCTTACGCTGTTGCTGCTGGTCCCAAGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTGGCCTCTAGAGGACAAAGCAGATACTTCAGGACCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACCAGAGCTCTTTGCAGAGATAGCATCCTGGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGACAAAGTCAAGAGAAACCACCATTCAATCCACCCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTACTTACCCACCGGGCAGGGAGCTCAAGAGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAATCTTCCACTCCAGCTTTGCAAAAGGACTCTTCTCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCTTACACTGTTTCGAGATGGCTTCCGCTTAAAAGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACTGTAAATAACCCCAAACTCCTGACCTCTAGAGGGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGCCAGGAAGGCGGTACCCACGCAGCTATGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAATGGAGCACTTGAGACAGGCTACCCCAGTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGGTGCTGCCCGAGCCAGTCTAATACAAAATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGTGGCCACAAAGGACATCGTGGGTCAGTTGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAACTTGTCATAATCACTGAGAAGCACCCAGTCCTTGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGGGGCATCAGTGAGAAAGCAAAGATGAACGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGATGGCTGCTGCTCACTCCTGTGCTGCTATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATGTTCAAAATGTTCAAACTGGTGAGGGCAAAGACCTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTCATCAAGGAAACCTGCAGCATAGCCTGGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTTCTAGAGCACACACTTGTCACTGACAGTTTGTCGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGAGGTGGTGCCAAGGGCCTCAAGTCTCCAATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCATGACTGTGTCTTTTCCACCTCTCTTCTCTGCGTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATTCCCACTTCAAAACAAGTGCGTACCATAACGTGACTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGTGGCTAGTTAGTCTTCAAGCCACAGAAGGAACAGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGTGGCCCAAAACATCTGAGAACTCTGCCCATTTCATTGCACCCTTGCACGATAC

HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCGGAATGGCCAAAGGCCATAGAGATCGTGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACTTGCTACTCTAGGTGCAGGGAAATATACTCCGAGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTGGCCGCCAGCATAGAAAGAGTAACGATTTCATGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACTCAGGCATGTGGCATTAGGATATTAGGATGTGCTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTATGAAAGATGCTGACAGAACTGGGCAACTAGCAGTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTCCTACAGCCTTCTTCGTGCCTGCCTTTCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCTCACCTCACACCAGGAGCAGCAAGCAGGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGGGCTGCTAATTCCTCAACTGTGCAGCTATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGAACTGTTGCTGCTCTTGGGGTCCATGGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCCCACTGGATCTGAGCTACCCCAGGACAGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTGGACAAGTCACCTCCAGCGTGACTCTCATTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCTGGGGTGTTCACTGAAGAAGAGCAAAGGTCTGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACGGAAATCAAGGAGATAGTCACCCTTCTCAGGTTCTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACCCCAATCTGTTGTTAACGTCTAGTCAAGTCAGCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCTTCCTTCTTAAATCAGTTTCCAAGGTGTTTCCGGGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCTTAAAGTCACTTCTCCACATTTGCTGTGGTATGGCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTACCCAGGAATCAGAGAAGGTAAAATGACCGAGACCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCATAGCCTGGAGATGTCTGTGACAGCTGAGGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATATGAATGTTCCTGCTCAGGGGCTCTAGGGACTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGGAAAGTGCTCTAATGAGTGAGCTCTGTGAGAACCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAACCAGCCAGGGAGGTGAGAATCAAATGGTACATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCAGGGAAAGGAGACACAGGAGTACACAACATTTAGGTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTTGGAAACACTTCCTCATGTCACTGTCTGAGGGTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAAGCTGGCTGGGAATCCCACCCACAACATTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCTAGTGGCCAGTCAAGGAGGGTGGCATACATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGAGAGGGACACTGGCCTCTCAGGAGGACCTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCAGCATCTGAATGGGTGACTGGAGTTATGGCAGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAAAGACATGATCTCTGGCGGCTGCGGCTGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTACCTGACGCTGCCTAGGATCACCAGAGGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTCTCTCATCGCTGAGCTGCAAGATGGCCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCTCGGGAGATATGGCTACCTACGGTGAAGATCCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTCGTGCCCAGGATTGCCAGGAGCCTAGAATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCCTGGGCAGGGAACTATGGCTAGGCAGGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTGGCCCCACATCCAAAGCCGCCCTTAAGGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCTTTGCAGCTAAGTGTTCGTGGAGAGAGGATGTGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTAATACCAGAGAATGAGGAAACCTTGCAAGCAGCCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTGCAGAGTAAAGTCCTAAGGATGGACTCTGCAGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCAGGATAAACCCTGGGAGGACTCTCTGGGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGGACTGAATGGAAGAAGCCGGGGAGACAGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGAGAGAAGACAGACACTGGCTACAAAGGTGTGTTTTCATTGCACCCTTGCACGATAC

HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTGTTCCTTTCCATAGGAGCTGGTGCTCACATCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACATGGCTCCCCATCTTCCAAGGACCTAGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAACATAAACAGTGAGGAACTCTTGTGGGTACACTGCGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGCAGTGGAGACAAGTGGGCCCTGTCCCTACTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCTCCCTTTCTAATAAAGCAACACTCATTGGCTCCCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGCTTGCCCATTTAATCTGTCAATCACCTTCACAGGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATCAAGATATTCACAACCCTTGCAGGACACACAGAGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTCCTGTACCTGTTTAATAAAGGGCCTCAAGACTCCCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGGTAGTGGAGAACTCCAATGGTCACAGAGAGAATGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCAGTGGCTTCTGCAGCTCTTGCTGGGTGGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGTGCTTTCTAAAGCCTGTGATCAATTGCTAGGCTCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTAGGGCATCAACAGGTGCTTCTTAGAGAAAAGACTGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAAAGACTGCCTACAGTTTTACAACAGCACGGGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACTGATCCTGCCTCCTGGACCCCAAGCAAGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTGAAGGGCAGAACTTTTGACATTCATGGGAGCACTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCCGCCCTCAGAGACCCCATATTGCTTATTTGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGGCAATGGAGAGTCCACTCTAACTTCCTGCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTCCACAGTTCTCAGAGTGGCATATTGGAACCAGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTAATACTGTGGGGTCCCCAGCTGTAACCACATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGCTATGCGTGCCCACTTTTCTAGTCTTCAGAAAAGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATTCTTGATACCAGGACTACCATCCTATCACTGGGCTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTGTAAAGACAGCACATTCTGTGGCCTTGACTGAAGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGCCACAAGAACTGGGGCACGCTGCTCCATCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATCCATGGAATATACCAGGTGTGTTCATTTCCACGCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTCTTCACAAAGCAAGAGTCACAACAGAGAAAGGTGCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCAAGAGCGCCTTGGGAAGGGGCATGAGGTATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTTCCAGACAGACACTGAGGTGGTACGAGTGAATACATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCACAGCAGCCCAGCTAACCTAAGATTATTATAGCCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCAGGCAGCTTTTCACTAATCACCAGGGCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAAGGCCAGTCTCACAGCAGGAAGGCAAAATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGACCCTCAGTCTAGGGGCACAGAAAGGTGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTTTGGTAAGGACCCAACTAAGGAGCTCCCTCTAGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACATTCCCTGAGAGTCCCCAAGCTGGTCCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGCATCGGTACACATGGCTTGCTGTGGACTATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCACTTCTCAGTTTATCTTCACACAGGAGGGGTAAGGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCACACTAGAGTAAAGCAGGAAGGCGGCAAACCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAGGTGACAATTGAAGAAGCAATCTGAAATGCTCGCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTTCCGCTTAAAACGCTTTGATGGGAAACCAAAGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGAAGGAAGCGAAAGGCAAGCAAGCGTTCAGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTGGACAAACTTGTCACTAATGGCAGTCTCTCCTACCTTTCATTGCACCCTTGCACGATAC

HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCAGAAGGCAGCAGGTCACCCAATTCCACCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGGGAGACACGTAGCATACATTTAACCTCATCTTGGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCAGGACAGAGAAGCCCACGTACAGAGAACTCTGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCCCAGCTAGCAGAACCAGTGAGGTACACTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAATTACACAGCACCCTCTATGCTCCAGGTTTTAGCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGTGGCATGTAACTCCCATACCCTCTTAAAAGCCCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAGTTCAGATGACTAACCCAGCTCTCTGCTCATGAAGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTAGCACTGACCACTGTGGGACATGACCTTTCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAGAGGACACTGATGCTCAACTACAACGGCTAAGTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCAAATCAAGGCTTTGCTGAATGGCCCTCTCCGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAACTGAACTCTACAATACGTGCTAAGTGCATGCCCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGTGAACTGTCAACACACACAAATGATTCCCTCCTGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTCTCTGCAGTTAGCTCTCTGCTCCAGTCTCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAATCTGGACTTCAGTCTACTGCATGTTAGGAAAGGCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTCAGAACAATCAACAAGATGAGGGATTCATGGCAAGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCTAGGTGTCTATGTGTTGGCCAGACCTCAGGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATTGAAGACAGATGGGCAGTCCAAAATACTCCAGACATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTGGATTCCTATATTGCTGTGTCCTGATCCCATCCAGATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTCTGCTTATTTAAAATGCTTACCCAACACGTGCCCTGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATAAGACAAAAGACAGACTGCACTTGTAGACCCCTGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAATTCAGACAAACATCAGGGATGGGTCTTAGGGAAGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTATGGGTCTACATTTATACAGCAAAGCCCATGAGAGCGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGAGCAGTCCTGGGTGCTGTTCCAATGGCCTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTACAGGTACTCTGGTGATGTTATACACAACCATACGGCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTTTGTACCGCTAATCAATTAGCTCGAACTTCCATGCCATTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTAAATTGGAAATCAGCAACTCAGGGATTCAAGGTCCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGGGGACAAATGCAGCATTGCAAATTCTAGGCCTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCACTGTTCTGGCCCAGTGCCAACAGGAAGGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGAGGGCTTCGTACCCTGGGGCGTCTCGTAGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGCCGCCGTCCCAGCCTACCTTCGCTATCAGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCCGCTTCATTAACGTCTGGAAGCGGCTACTTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTGCTTTCCTTAAGCTGCTGATGCAGCCTGCGTTTCATTGCACCCTTGCACGATAC
HJURP_pool	TGAATAGCAGCGGTGGCAAACTTTCACCCGCCCAAGCCCGTCCACAGATCAAACTTTCATTGCACCCTTGCACGATAC