DRAFT Report to the G10K genome assembly workshop on the results of the Assemblathon contest DRAFT

Dent Earl Benedict Paten John St. John Keith Bradnam
Joseph Fass Dawei Lin Aaron Darling Ngan Nguyen
Mark Diekhans Ian Korf David Haussler

March 15, 2011

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Participants

There were 39 assemblies submitted by a total of 17 teams, all of which are listed in Table 1.1. Each assembly is identified by the ID letter of the team that submitted it followed by the within team number of the assembly. ID Letters V and beyond are assemblies generated by the UC Davis analysis team in an attempt to provide a more rich distribution of predictions.

1.1.1 Assembly Details

Information provided by assembly teams is displayed in this section.

A1

Software: PE-Assembler

Main Focus: Use up all reads (include them in contigs)

Computational Requirements: Used 64 cores in a SMP, ∼30GB memory and ∼6-8 hours runtime.

Used Reference: No.

B1

Software: Phusion2, phrap.

Main Focus: Maximixe N50 with good accuracy.

Computational Requirements:

1. Read pre-assembly process 1 CPU hour single processor 5GB RAM
2. Reads clustering 3 CPU hours single processor 100GB RAM
3. Contig generation 50 cores each for 0.5 hours with 4GB RAM
4. Supercontigs 2 CPU Hours single processor 20GB RAM

Used Reference: No.

Table 1.1: Groups that submitted assemblies to the contest.

ID	Team name	Affiliations	Principal contact	Entries	Software
A	GIS_CMB1	Agency for Science, Technology and Research, Singapore	Pramila Ariyaratne	1	PE-Assembler
B	Phusion	Wellcome Trust Sanger Institute, UK	Zemin Ning	2	Phusion2, phrap
C	Ensembl Genomes' Curtain	European Bioinformatics Institute, UK	Matthias Haimel	2	SGA, BWA, Curtain, Velvet
D	sanger-sga	Wellcome Trust Sanger Insitute, UK	Jared Simpson	4	SGA
E	Borgs	CRACS (Center for Research in Advanced Computing Systems), Portugal	Nuno Fonseca	3	ABySS
F	ABySS	BC Cancer Genome Sciences Centre, Canada	Shaun Jackman	5	ABySS, Anchor
G	Plant Genome Assembly Group	DOE Joint Genome Insititute, USA	Jarrod Chapman	1	Meraculous
H	Team Symbiose	L'IRISA (Institut de recherche en informatique et systèmes aléatoires), France	Rayan Chikhi	5	Monument
I	Terrapins	CSHL (Cold Spring Harbor Laboratory), USA	Michael Schatz	2	Quake, Celera, Bambus2
J		Department of Computer Science, Iowa State University	Xiaoqiu Huang	1	PCAP
K	Super Dawgs	Computational Systems Biology Laboratory, University of Georgia, USA	Wen-Chi Chou	3	Seqclean, SOAPdenovo
L	PRICE @ deRisi Lab	UC San Francicso, USA	Graham Ruby	1	PRICE
M	Softberry	Royal Holloway, University of London, UK	Victor Solovyev	5	OligoZip
N	TGAC/TSL/Oxfor	The Genome Analysis Centre, Sainsbury Laboratory, and Wellcome Trust Centre for Human Genetics, UK	Mario Caccamo	3	Cortex_con_rp
O	KAS	Department of Computer Science, University of Chicago, USA	Fangfang Xia	1	Kiki
P	BGI-Shenzhen	BGI, China	Zhenyu Li	1	SOAPdenovo
Q	ALLPATHS Assembly Team	Broad Institute	David Jaffe	1	ALLPATHS-LG
V	UCD	—	—	6	Velvet
W	UCD	—	—	12	CLC
X	UCD	—	—	6	ABySS

B2

Software: Phusion2, phrap, SOAPdenovo
Main Focus: Same as B1.
Computational Requirements: Same as B1.
Used Reference: No.

C1

Software: SGA (unpaired), BWA, Curtain (uses Velvet)
Main Focus: Minimized contamination with improved N50 length.
Computational Requirements: SGA:

- Max memory: ∼6GB
- Total cpu time ∼290 CPU hours (submitted to a cluster)

BWA:

- Max memory: ∼0.5 GB
- Total cpu time: ∼62 CPU hours (submitted to a cluster)

Curtain pipeline:

- Max memory: ∼ 16 GB
- Total cpu time: ∼ 9 CPU hours (submitted to a cluster)

Combined:

- Max memory: ∼ 16 GB
- Total cpu time: ∼ 361 CPU hours (submitted to a cluster)

Used Reference:

C2

Software: Velvet.
Main Focus: Focused to reduce contamination and improve the N50 length.
Computational Requirements:
Used Reference: No.

D1

Software: The assembly has performed with sga (string graph assembler). All the code for the assembler can be found at: `https://github.com/jts/sga` . This was the only program used except for the DistanceEst stage of the ABySS assembler, which we used to estimate the distances between contigs from paired end data.

Main Focus: This is assembly 1 of 4. As we found significant structural variation in the diploid genome, we made assemblies with various levels of stringency.

In this assembly, we aggressively attempted to resolve structural variation and incorporate one of the variants into the scaffolds. The scaffolds will switch between haplotypes in places. No attempt to resolve the gap between contigs in the scaffolds was made.

Computational Requirements: sga is a multi-staged assembler. For most stages, the input data can be broken up to run in parallel on a cluster. We used up to 16 jobs for the longest stage (correcting base calling errors in the reads). The aggregate CPU time (summing all processes and threads) was approximately 300 hours. The maximum amount of memory used by a single process was 5.8GB.

Used Reference: No.

D2

Software: Same as D1.

Main Focus: This is assembly 2 of 4. As we found significant structural variation in the diploid genome, we made assemblies with various levels of stringency.

In this assembly, we aggressively attempted to resolve structural variation and incorporate one of the variants into the scaffolds. The scaffolds will switch between haplotypes in places. In this build, we attempted to fill in the gaps between scaffold contigs by searching for a walk through the assembly graph connecting the contigs.

Computational Requirements: Same as D1.

Used Reference: No.

D3

Software: Same as D1.

Main Focus: This is assembly 3 of 4. As we found significant structural variation in the diploid genome, we made assemblies with various levels of stringency.

In this assembly, we broke the scaffolds when we found potential structural variation. This is a conservative set of scaffolds. No attempt to resolve the gap between contigs in the scaffolds was made.

Computational Requirements: Same as D1.

Used Reference: No.

D4

Software: Same as D1.

Main Focus: This is assembly 4 of 4. This is the set of contigs that were constructed from the 200 and 300bp PE libraries. These contigs were constructed from the read sequences alone without using the pairing information.

Computational Requirements: Same as D1.

Used Reference: No.

E1

Software: Assembly pipeline using Abyss for contig generation.

Main Focus:

Computational Requirements: Six-Core AMD Opteron(tm) processor (800MHz) with 128 GB of RAM, and running Fedora Core 13. The approximate total amount of computation time required to generate the assembly was 345 hours.

Used Reference: Yes, only for ordereing of scaffolds after they had been built.

E2

Software: Same as E1.

Main Focus:

Computational Requirements: Same as E1.

Used Reference: Yes, only for ordereing of scaffolds after they had been built.

E3

Software: Same as E1

Main Focus:

Computational Requirements: Same as E1.

Used Reference: Yes, only for ordereing of scaffolds after they had been built.

F1

Software: ABySS (version 1.2.6).

Main Focus: "Vanilla" ABySS assembly, using default options of the most-recent released version.

Computational Requirements: 40 GB of RAM.

Used Reference:

F2

Software: ABySS (version 1.2.6) and Anchor.

Main Focus: Uses only paired-end data, no mate-pair data.

Computational Requirements: Same as F1.

Used Reference:

F3

Software: ABySS (version 1.2.6) and Anchor.

Main Focus: Uses the paired-end data and the forward-reverse oriented reads of the mate-pair data (500 bp fragments), but not the larger reverse-forward oriented reads.

Computational Requirements: Same as F1.

Used Reference:

F4

Software: ABySS, Anchor and an experimental scaffolding algorithm.

Main Focus: An experimental scaffolding algorithm; based on assembly #2.

Computational Requirements: Same as F1.

Used Reference:

F5

Software: ABySS, Anchor and an experimental scaffolding algorithm.

Main Focus: An experimental scaffolding algorithm; based on assembly #3.

Computational Requirements: Same as F1.

Used Reference:

G1

Software: Meraculous (Chapman et al., submitted to PLOS One).

Main Focus: This assemblathon submission is our first attempt at navigating the deBruijn graph through SNP/indel polymorphisms in a diploid genome. At polymorphic loci contained within the contigs, the assembly represents the allele that is most prevalent in the simulated dataset (the "max-depth" allele). The identity of the alternate allele is also known, but not represented in the submitted FASTA. The present assembly does not attempt to capture information about long-range haplotypes. We did not make any explicit attempt to remove any "contamination" but rather to accurately assemble the reads in the dataset as completely as possible.

Computational Requirements: Meraculous has several stages, most of which can be distributed on commodity clusters. For the Assemblathon dataset, the Initial distributed mer-steps took about 1 hour on 256 cores each with access to ∼4 GB RAM. Generating contigs is fast (minutes) and requires minimal RAM (¡2 GB) on a single core. Each round of scaffold-building "order and orientation" takes ∼45 minutes to map the reads to the contigs under conditions that guarantee uniqueness with blastMap (using ∼50 cores). An optional gap-closing step is used with standard paired-ends. This required 1 hour on a single core. The total wall clock time for the reported assemblathon assembly was around 6 hours.

Used Reference: No.

H1

Software: Indexing, assembly of paired-end reads: Monument. Assembler: new, experimental assembler written mainly in Python. It is based on paired string graphs; a manuscript describing the algorithm was submitted to JCB. Post-processing: MUMmer 3.0.

Main Focus: Main focus: show feasibility of parallel, low-memory assembly. Maximize N50 length of contigs constructed from paired-end (200bp and 300bp) libraries. Mate-pairs were not used. We are sending the raw assembly produced by Monument and filtered with MUMmer, no verification of manual finishing was performed, as well as no preliminary error-correction.

Computational Requirements: On a "desktop" computer using 1 thread: ∼ 2 days and 20 GB of RAM.

On a 7 nodes cluster (totaling 70 cores), the main assembly phase takes between 50 and 90 minutes – depending on parameters (80 minutes for this specific assembly).

I'd like to make a late update about Monument assembler computational resource usage. We optimized it for the dngasp competition, it can now assemble the Assemblathon paired-end dataset with strictly the same index, compressed down to

6.3 GB, in 2 hours (got rid of the MUMmer step). Not sure if you're going to use this information, but I thought I would mention it :)

Used Reference: No.

H2

Software: Same as H1 plus SSPACE (`http://www.baseclear.com/sequencing/data-analysis/bioinf` using all libraries.

Main Focus: scaffolding of Assembly H2, using an external scaffolding tool (SSPACE with default parameters, scaffolding mode, i.e. no contig extension).

Computational Requirements: Same as H1 plus SSPACE runs in 35 minutes on a single node with negligible memory usage.

Used Reference: No.

H3

Software: Same as H1.

Main Focus: We changed one parameter (consensus threshold from 2 to 3) from the previous submission (Assembly H1) and it appears to produce a better assembly (still with paired-end libraries only).

Computational Requirements: Same as H1.

Used Reference: No.

H4

Software: Same as H1.

Main Focus: These are two assemblies we obtained using an earlier, less strict version of Monument using only paired-end reads. Mate-pairs were not used; no scaffolding was done. We expect them to have a significantly higher misassembly rate. Highest N50: monument_pairedend_nonstrict.fa.bz2 Longest contig: monument_pairedend_nonstrict.fa2.bz2

Computational Requirements: Same as H1.

Used Reference: No.

H5

Software: Same as H1.

Main Focus: Same as H4.

Computational Requirements: Same as H1.

Used Reference: No.

I1

Software: Quake (`http://www.cbcb.umd.edu/software/quake/`) - Pre-assembly error correction and trimming. Celera Assembler (`http://wgs-assembler.sf.net`) - Primary Assembler. Bambus2 (`http://amos.sf.net/bambus2`) - Scaffold Refinement.

Main Focus: Maximize scaffold span - it is possible/likely that we merged the haplotypes in the process.

Computational Requirements: Quake - roughly 1 day on 24 cores, peak memory around 100GB. CA - roughly 2 days on 24 cores, peak memory around 100GB. Bambus2 - roughly 2 days on 1 core, peak memory around 100GB.

Used Reference: No.

I2

Software: Quake (`http://www.cbcb.umd.edu/software/quake/`) - Pre-assembly error correction and trimming. Celera Assembler (`http://wgs-assembler.sf.net`) - Primary Assembler.

Main Focus: Maximize overall sequence composition.

Computational Requirements: Quake - roughly 1 day on 24 cores, peak memory around 100GB. CA - roughly 2 days on 24 cores, peak memory around 100GB.

Used Reference: Yes.

J1

Software: PCAP package for solexa.

Main Focus: De novo assembly.

Computational Requirements: It depends, we used 400 cores for assembling this genome.

Used Reference: No.

K1

Software: Seqclean Correction tool for SOAPdenovo. SOAPdenovo. GapCloser for SOAPdenovo.

Main Focus: Maximize N50.

Computational Requirements: 8 cores with total 128G RAM.

Used Reference:

K2

Software: Seqclean Correction tool for SOAPdenovo. SOAPdenovo. GapCloser for SOAPdenovo. GMAP. blastall.

Main Focus:

Computational Requirements: Same as K1.

Used Reference:

K3

Software: Same as K2.

Main Focus: minimize bacterial contamination.

Computational Requirements: Same as K1.

Used Reference:

L1

Software: PRICE is designed to assemble viral genomes from metagenomic datasets.
Main Focus:
Computational Requirements:
Used Reference:

M1

Software: OligoZip.
Extra Info: 650 set is one iteration accounting 3000 and 1000 read pairs with 5000 set). softberry_assembly_658.fa.gz
Computational Requirements: 10 processors, 16 Gb.
Used Reference: No.

M2

Software: Same as M1.
Extra Info: 5000 set is a little processed biggest contigs from initial set (we remove relatively small ones as we had not enough time to make some further steps (accounting pairs) with full set above) - softberry_assembly_5000.fa.gz
Computational Requirements:
Used Reference: No.

M3

Software: Same as M1.
Extra Info: The ∼40 000 is the initial clustering set (not accounting for 10000 and 3000 pairs) - softberry_assembly_40344.fa.gz
Computational Requirements:
Used Reference: No.

M4

Software: Same as M1.
Extra Info:
Computational Requirements:
Used Reference: No.

M5

Software: Same as M1.
Extra Info:
Computational Requirements:
Used Reference: No.

N1

Software: cortex_con_rp. cortex_con_rp is new software development and still work in progress. This is a collaboration between the Bioinformatics division at TGAC (Norwich), Bioinformatics department at The Sainsbury Lab (Norwich) and the department of Statistics in Oxford University.

Main Focus: High performance. Low memory usage (suitable for large mammalian or plant genomes). Accuracy. Completeness.

Computational Requirements: One 64Gb machine (single threaded) – it takes under 24hs. In a cluster this can be reduced to under 15 hs.

Used Reference: No.

N2

Software: Same as N1.

Main Focus: Low memory usage, fast processing, high accuracy of contigs.

Computational Requirements: Initial merge of read files required 64Gb of RAM. After graph cleaning, later steps required only 16Gb of RAM. Using single threading, 1 core, and in series, time taken from start to finish approx. 32 hours. Carrying out some steps in parallel on a cluster (still single threaded), process complete in approx. 24 hours.

Used Reference: No.

N3

Software: Same as N1.

Main Focus: Low memory usage, fast processing, high accuracy of contigs.

Computational Requirements: Same as N2.

Used Reference: No.

O1

Software: Kiki

Main Focus: Parallel performance.

Computational Requirements:

Used Reference:

P1

Software: SOAPdenovo. The SOAPdenovo is the lastest beta version and still under test.

Main Focus: maximize N50 length and minimize bacterial contamination.

Computational Requirements: 16 cores, 70GB RAM, run time is not calculated, it should be less than a day.

Used Reference: No.

Q1

Software: ALLPATHS-LG.
Main Focus: Best assembly possible.
Computational Requirements: Requires 100 GB RAM to run. Run time was 12 hours on a 48 core server.
Used Reference: No.

V1

Software: Velvet.
Main Focus: untrimmed reads, all read libraries, all pairing info.
Computational Requirements:
Used Reference: No.

V2

Software: Same as V1.
Main Focus: untrimmed reads, all read libraries, no MP info.
Computational Requirements:
Used Reference: No.

V3

Software: Same as V1
Main Focus: untrimmed reads, all read libraries, no MP or PE info
Computational Requirements:
Used Reference: No.

V4

Software: Velvet.
Main Focus: trimmed reads, all read libraries, all pairing info
Computational Requirements:
Used Reference: No.

V5

Software: Velvet.
Main Focus: trimmed reads, all read libraries, no MP info.
Computational Requirements:
Used Reference: No.

V6

Software: Velvet.
Main Focus: trimmed reads, all read libraries, no MP or PE info.
Computational Requirements:
Used Reference: No.

W3

Software: CLC
Main Focus: untrimmed reads, all read libraries, no MP or PE info.
Computational Requirements:
Used Reference: No.

W5

Software: CLC
Main Focus: trimmed reads, all read libraries, no MP info
Computational Requirements:
Used Reference: No.

W6

Software: CLC
Main Focus: trimmed reads, all read libraries, no MP or PE info
Computational Requirements:
Used Reference: No.

W7

Software: CLC
Main Focus: untrimmed reads, all read libraries, F/F orientation for mate pairs (CLC only)
Computational Requirements:
Used Reference: No.

W8

Software: CLC
Main Focus: untrimmed reads, all read libraries, F/R orientation for mate pairs (CLC only)
Computational Requirements:
Used Reference: No.

W9

Software: CLC
Main Focus: untrimmed reads, 300i paired, all others as single (CLC only)
Computational Requirements:
Used Reference: No.

W10

Software: CLC
Main Focus: untrimmed reads, 300i only, no pairing info (CLC only)
Computational Requirements:
Used Reference: No.

W11

Software: CLC
Main Focus: trimmed reads, all read libraries, F/F orientation for mate pairs (CLC only)
Computational Requirements:
Used Reference: No.

X1

Software: ABySS
Main Focus: untrimmed reads, all read libraries, all pairing info.
Computational Requirements:
Used Reference: No.

X2

Software: ABySS
Main Focus: untrimmed reads, all read libraries, no MP info
Computational Requirements:
Used Reference: No.

X3

Software: ABySS
Main Focus: untrimmed reads, all read libraries, no MP or PE info
Computational Requirements:
Used Reference: No.

X4

Software: ABySS
Main Focus: trimmed reads, all read libraries, all pairing info
Computational Requirements:
Used Reference: No.

X5

Software: ABySS
Main Focus: trimmed reads, all read libraries, no MP info
Computational Requirements:
Used Reference: No.

X6

Software: ABySS
Main Focus: trimmed reads, all read libraries, no MP or PE info
Computational Requirements:
Used Reference: No.

Chapter 2

Methods

2.1 Genome simulation

To simulate the effects of evolution on a genome we used the EVOLVER suite of genome evolution tools provided by Edgar, Asimenos, Batzoglou and Sidow [Edgar et al., 2010]. EVOLVER is capable of simulating the forward–time evolution of multi–chromosome haploid genomes and includes models for evolutionary constraint, protein codons, genes and mobile elements. The simulation was managed by a set of scripts (to be released shortly) that control the execution of EVOLVER programs and scripts, and manage the direction and storage of input and out files and parameters. EVOLVER opperates with a concept of time the authors have termed "ticks," and which are roughly analogous to the number of neutral substitutions per site.

The input, or "root," genome for the simulation was constructed by downloading the DNA sequence for human chromosome 13 (hg18/NCBI36) and the annotations UCSC Genes, UCSC Old Genes, CpG Islands, Ensembl Genes and MGC Genes from the UCSC table browser[Fujita et al., 2010]. This "genome" was then divided into four chromosomes of approximately the same size. The genome was then coupled with parameters and a mobile element library provided by Arend Sidow (pers. comm.) to form the EVOLVER infile set for the simulation.

Figure 2.1 shows the phylogeny used to generate the simulated genomes. We used an EVOLVER step size of 0.01, meaning the initial branch length of 0.40 ($\sim$ 200 my) from the root node to the most recent common ancestor (MRCA) of the final leaf genomes node consisted of 40 separate EVOLVER cycles. We performed this long burnin on the genome in an attempt to reshuffle the sequence and annotations present in the genome to prevent naive discovery of the source of the root genome. The A and B clades are symmetrical with respect to the phylogeny. The lineage leading to A descends from the MRCA for a distance of 0.1 ($\sim$ 50 my) and then splits into the lineages leading to the leaf genomes A1 and A1, which are 0.002 ($\sim$ 1 my) divergent from one another.

We released the sequence of of the B1, B2, and B genomes and their annotations to the community for use by any assemblers capable of using comparative phylogenetic methods. Table 2.1 provides a count of some of the events that took place during the simulation and times their occurence to particular branches in the

Figure 2.1: Phylogeny showing the evolution of the simulated genomes. The root genome, chromosome 13 from human (hg18/NCBI36) and split into quarters each approximately 28 Mb, is shown on the left. This lineage was allowed to evolve for a distance of 0.40 neutral substitutions per site (NSS), or approximately 200 my. The resulting genome is the most recent common ancestor (MRCA) for the A and B genomes. The A and B lineages are symmetric with respect to the phylogeny. The A lineage was evolved for a distance of 0.1 NSS, or 50 my. This lineage was then split into two, A1 and A2, which were evolved for a distance of 0.002 NSS, or 1 my.

phylogenetic tree.

Sequences A1 and A2 were used as input to the sequence simulation aspect of the project, covered in section 2.2.

2.2 Sequencing simulation

2.2.1 High Level Overview

We needed a program that would generate short reads, model at least some of the most glaring sources of error the mate-pair protocol introduces, have ambiguous read IDs and also maintain the location information about where the reads originate. We also didn't want or need the program to model evolution on the underlying

Table 2.1: Number of Evolver events timed to branch lineages. The total number of events between the root and genome A1 would thus be the sum of the burnin events, the A events, and the A1 events.

	Burnin	A	B	A1	A2	B1	B2
Substitutions	35,900,803	9,702,486	9,710,428	197,117	196,896	196,872	196,803
Deletions	2,465,209	672,740	674,273	13,528	13,834	13,632	13,621
Inversions	11,701	3,321	3,313	54	61	64	71
Moves	4,714	1,369	1,325	34	31	26	35
Copies	14,644	4,151	4,043	83	80	82	79
Tandem	1,161,841	313,255	313,958	6,436	6,494	6,354	6,445
Chr Split	2	1	0	0	0	0	0
Chr Fuse	4	0	0	0	0	0	0
Create CDS	44	6	7	1	0	0	0
Delete CDS	30	10	9	1	0	0	0
Create UTR	4	2	1	0	0	0	0
Delete UTR	2	1	1	0	0	0	0

sequence, since that is something already taken care of by EVOLVER. Since all of these features were not available in existing software, we wrote our own short read simulator we are calling SimSeq (`http://github.com/jstjohn/SimSeq`).

At a very high level, reads are either sampled from the genome with a paired-end strategy, or a mate-pair strategy, and then we apply an error profile to those reads in their proper orientation (facing inwards for paired-end, and outwards for mate-pair).

2.2.2 Read Sampling Strategy

- For read sampling we employed two separate methods, one for mate-pair libraries and the other for paired end libraries.
- Reads are sampled uniformly across each sequence.
- Coverage depth is kept approximately uniform by weighting the number of reads sampled from each sequence by its length.
- Read fragments may be sampled from either strand with equal probability
- Duplicates are produced with some probability before the error is applied to the reads.

Paired-End Sampling

Illumina paired end sampling is the most straightforward strategy which simply involves fragmenting the genome and size selecting for pieces typically in the 150-500 base pair range. We allow the user to select any mean level they want along with a standard deviation. The fragment size is then sampled with a normal distribution. The reads are oriented facing each other and may be sampled from either strand.

Mate-Pair Sampling

Illumina mate-pair library construction differs from paired-end library construction in important ways that can introduce several types of important error into reads. A high level overview of how it works is to attach a chemical tag onto the ends of a long fragment, typically in the range of 2-10kb, then circularize the fragment. After the fragment is circularized, it is fragmented into sizes typically in the 200-500 base range, and it is enriched for fragments that contain the chemical marker which is located at the region where the loop is joined.

There are three common types of error introduced in the mate-pair library prep process, and we currently model two of them. First off when the loop is formed, there is a chance that a loop will be formed between two non-related long fragments, and the reads would be from unrelated parts of the genome in that case. We do not model that type of error. Assuming that the loop has been formed properly, one type of error results from a fragment that doesn't contain the chemical tag mistakenly being sampled. This results in a paired end read with an insert size of whatever loop fragmentation protocol insert size was chosen to be present in the library. We do model that type of error. The next major source of error is from the

random fragmentation process resulting in the loop join position occurring in the middle of a read rather than between the two reads. We model this by assuming a uniform distribution of loop join sites across a sampled loop fragment, which results in chimeric reads as a function of the size of the fragmented loop piece, and the length of the reads. For example shorter reads and longer loop fragmentation pieces are less likely to result in a chimeric read.

2.2.3 Base-Level Error Model

For Assemblathon 1 we utilized an error model that is dependent on the position within the read and the underlying reference base. To generate this model we assembled a human mitochondrial genome using reads from an Illumina HiSeq run with the reference guided assembler MIA (`http://sourceforge.net/projects/mia-assembler/`). We then took that assembly and mapped all reads back to it using BWA (`http://bio-bwa.sourceforge.ne` with default settings to do a paired end mapping to the sequence. We kept all alignments with a mapq quality score over 10. We then iterated through the alignment and built an empirical distribution of phred scores and probabilities of observing one of [A,C,G,T,N] given the reference base and position in the read at each phred score. In this way we have an error model that is conditioned on the phred score, position, and reference base, but does not make any assumptions about the actual underlying error of each phred score at each position.

2.2.4 SimSeq Settings used for Assemblathon 1

In addition to the diploid genome pool, I added 3 copies of the E. coli sequence (gi 312944605) to yield a 5% E. coli contamination rate. We later realized that the E. coli reference sequence has several degenerate bases, so those ended up in some of our reads.

Paired-End Libraries

1. 200bp insert +/- 20 standard deviation.
   (a) 2x 100bp
   (b) 22499731 read pairs (40x coverage of the diploid sequence)
   (c) 0.01 probability of being a duplicate
2. 300bp insert +/- 30 standard deviation.
   (a) 2x 100bp
   (b) 22499731 read pairs (40x coverage)
   (c) 0.01 probability of being a duplicate

Mate-pair libraries

1. 3000bp loop length +/- 300 standard deviation
   (a) 2x 100bp

(b) 500 bp loop fragmentation size +/- 50 bp

(c) 0.2 probability of sampling a PE fragment rather than an MP fragment

(d) 11249866 read pairs (20x coverage)

(e) 0.05 probability of being a duplicate

2. 10000bp loop length +/- 1000 standard deviation

   (a) 2x 100bp

   (b) 500 bp loop fragmentation size +/- 50 bp

   (c) 0.3 probability of sampling a PE fragment rather than an MP fragment

   (d) 11249866 read pairs (20x coverage)

   (e) 0.08 probability of being a duplicate

Problems With the Error Model Used in Assemblathon 1

The first problem with this strategy is that we end up leaving out most reads that have an error rate that is too high to confidently map to our assembled mitochondria. We could partially get around this by using the PhiX control lane and do a more sensitive mapping back to the PhiX 174 genome where we can try much harder to place each read.

The next problem with our strategy is that since we do not condition on the prior phred score you will notice a mixture of great and poor quality bases at the ends of the reads. Since each position is independent of the previous one, you don't see the easy to trim string of bad phred scores that one typically sees in a real data set.

The final problem was due to human error in generating the error model. During development there was a bug that resulted in a reversed error model where the first position was actually the last position in the reads and so on. This problem was identified and fixed early on in development before the Assemblation, however the reversed error model was mistakenly used to generate the Assemblathon 1 data. This resulted in the bases with a slightly higher error rate tending to appear towards the beginning of the reads rather than towards the end of reads (see Figure 2.2). This reversed error model was also reflected in the phred score, so as long as people removed bases with bad phred scores from both ends of the reads rather than just from the end of reads this wouldn't have been a major issue. In either case the error rate for Assemblathon 1 was fairly low due to the error sampling strategy being biased towards reads with lower error rates.

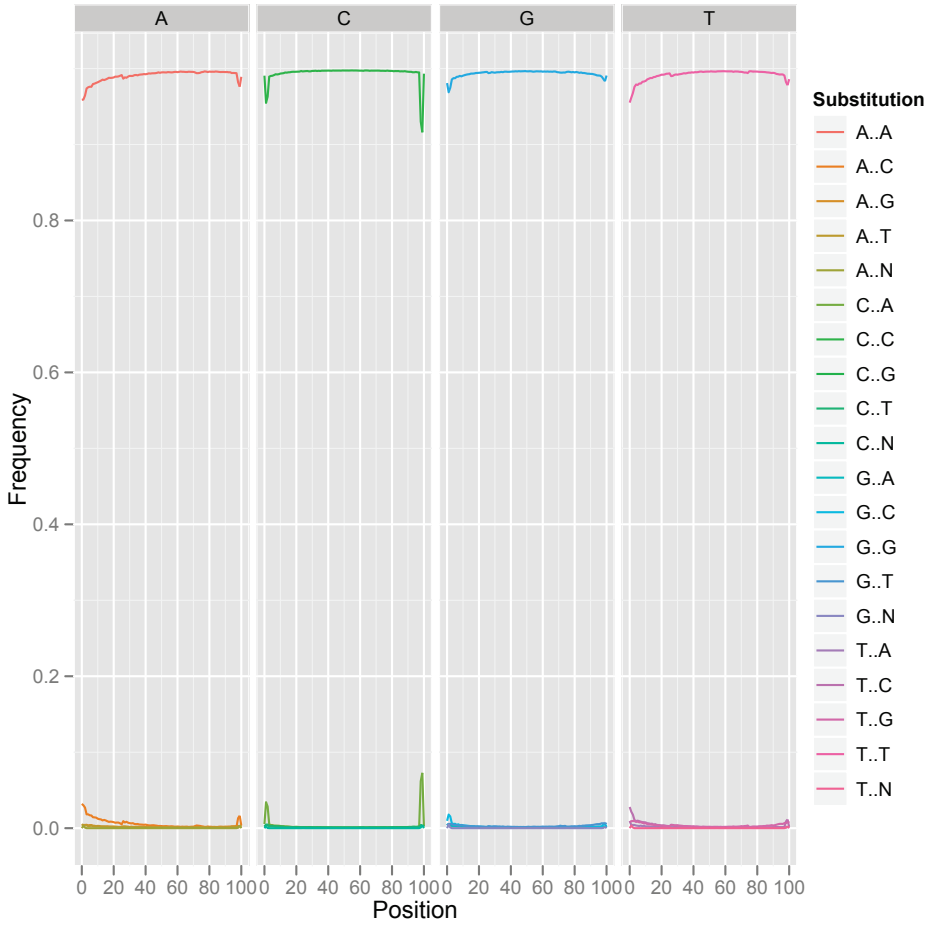


Figure 2.2: Reference base and read-position specific nucleotide sampling probabilities averaged over all phred scores. The line near the top is the correct base, and the three lines toward the bottom are the three alternate non-reference bases. Notice that the error rate is slightly worse at the beginning of the read. The correct error model should have been in reverse. This error model was generated from a high coverage mitochondrial alignment of 2x100bp paired-end Illumina HiSeq reads.

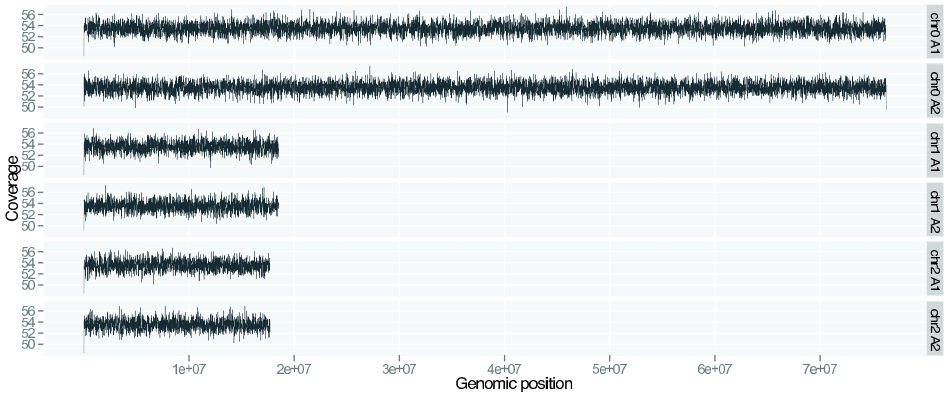


Figure 2.3: Coverage of simulated sequencing reads.

2.3 Assembly analysis

2.3.1 Masking

Every assembly, each haplotype and the bacterial contamination were soft masked for repeative elements using both Tandem Repeats Finder [Smit, 1999] and RepeatMasker [Benson et al., 1999], following the standard UCSC genome browser comparative genomics pipeline [Fujita et al., 2010].

2.3.2 Cactus Alignment Generation

To compare an assembly with the simulated genome and bacterial contamination we constructed a multiple sequence alignment (MSA). The sequence inputs to the MSA were the two haplotypes, each with three chromosomes, the bacterial contamination, a single chromosome, and the sequences of the assembly. To generate the MSA we used an adapted version of the newly developed Cactus [Paten et al.,] alignment program, a new MSA program able to handle rearrangements, copy number changes (duplications) and missing data.

The Cactus program starts by using the Lastz (`http://www.bx.psu.edu/miller_lab/dist/README.la` pairwise alignment program to generate a set of pairwise alignments between all the sequences and themselves. In the adapted version of Cactus used for the assemblathon, which we hence forth call Cactus-A, we used the following parameters to Lastz, after discussion with the program's author: '–step=10 –seed=match12 –notransition –mismatch=2,100 –match=1,5 –ambiguous=iupac –nogapped –identity=98'. This ensured that the resulting pairwise alignments were ungapped (without insertions and deletions), of minimum length 100 and with an identity (sequence similarity) of 98% or greater. Cactus-A uses these alignments to build a "sparse map" of the homologies between its input sequences. Once this sparse map is constructed, in the form of a Cactus graph [Paten et al., 2011], a novel algorithm is used to align together sequences that were initially unaligned in the sparse map. To prevent sequences that are not homologous from being aligned in this process we set the alignment rejection parameter, called γ, in Cactus-A to 0.5, to filter positions from being aligned that are not likely to have very recently been diverged. The result of the adaptations in Cactus-A is a high specificity map of the alignment of the assembly to the two haplotypes and the bacterial contamination. It should be noted that in the future we will be able to use the simulated evolutionary relationships between the two simulated haplotypes, rather than rediscovering this relationship in the MSA, but, due to the technical difficulties of this, we chose to realign these sequences together. The results of Cactus-A are stored as MAF file [Blanchette et al., 2004], these will shortly be available for download. As we aligned both the bacterial contamination and the two haplotypes together we used the hypothetical existence of any alignments between the haplotypes and the bacterial contamination as a negative control for non-specific alignment. We did not observe any such alignments.

2.3.3 Blocks, Haplotype Paths and Scaffold Paths

A MSA can be described as a graph, we call the simplest such graph that we use an *adjacency graph*. A formal description of the adjacency graph can be found in [Paten et al., 2011], it is closely related to a bi-directed bigraph representation of a de-Bruijn used in assembly [Medvedev and Brudno, 2009] [Zerbino and Birney, 2008] and the multi breakpoint graph used in the study of genome rearrangements [Alekseyev and Pevzner, 2007]. The graph contains two kinds of edges, *block edges*, which represent gapless (containing no-indels or structural rearrangements of any kind) maximal blocks of alignment, and *adjacency edges*, which represent collections of connections between the ends of segments of DNA. The nodes in the graph represent the ends of blocks of aligned sequences. Figure 2.4 illustrates an example.

Here we informally define a *block* as a maximal gapless alignment of a set of *homologous* sequences, each block is therefore represented as a block edge in the adjacency graph. For a formal definition see [Paten et al., 2011]. A block can be divided up into *columns*, each of which represents a set of individual base pair positions in the input sequences that are considered homologous. The base pair length is equal to the number of columns that it contains. The block structure of an MSA is defined by the discovered homology relationships between the sequences. Sequences that are very closely related are likely to exhibit fewer blocks with a higher base pair length than sequences that are significantly diverged from one another. The two haplotypes are sufficiently polymorphic with respect to one another that the median base pair length of blocks in an alignment of just the two haplotypes is around 4000 bases. As this length is much less than the length of many sequences in the assemblies, assessing an assembly requires methods that do not penalise when the assembly reconstructs a polymorphism, to this end we construct a graph theoretic analysis.

Within the adjacency graph a sequence is represented as a path of alternating adjacency and block edges, termed a *thread*. We can assess the accuracy of assembly sequences by analysing their thread representation in the adjacency graph. Let P be the thread representing an assembled sequence in the adjacency graph, G. Each edge in G is labelled with the sub-sequences it represents, called *segments*, thus it is possible to discern if the edge represents segments in the haplotypes, the assembly and/or the bacterial contamination. As previously stated, no edges are contained in G which represent segments in both the haplotypes and the bacterial contamination. Any edge e in P is *consistent* if that edge is also labelled with segments from either or both of the haplotypes. For any P a *haplotype path* is a maximal subpath of P in which all the edges are consistent. Thus P can be divided up into a series of haplotype paths, possibly interspersed with edges in P that are not contained in a haplotype path, see Figure 2.5 for an example. The *base pair length* (or simply length, when it is unambiguous) of a haplotype path is equal to the sum of the base pair lengths of the block edges it contains. Haplotype paths represent maximal portions of the assembled sequence which are consistent with one or both of the haplotypes.

Many of the assembled sequences contain scaffold gaps, where the assembly program has inferred two individually assembled sequences are contiguous, to account for these gaps we can define scaffold paths. Again, let P be a thread representing

an assembled sequence in the adjacency graph. Any block edge e in P is *ambiguous* if the segment of assembly sequence that labels it contains wildcard characters (denoted as 'N's).

Let (a, b) and (c, d) be two block edges in P and (b, c) and (d, e) be two adjacency edges in P, such that (a, b) represents the end of one haplotype path and (c, d) represents the start of another. We call this structure a *scaffold gap* if (1) either b or c is ambiguous and (2) there is a path of edges labelled with haplotype sequences connecting b to c. This structure is illustrated in Figure 2.6(a). It is also possible that the wild card characters create an insertion, in this case let (a, b) and (c, d) and (e, f) be three block edges in P and (b, c) and (d, e) be two adjacency edges in P, such that (a, b) represents the end of one haplotype path and (e, f) represents the start of another. We call this structure a *scaffold gap* if (1) either (c, d) is ambiguous and (2) there is a path of edges labelled with haplotype sequences connecting b to e. This structure is illustrated in Figure 2.6(b). Making this two case definition allows for the wild card chracters to be aligned or unaligned, and therefore makes the definitions of scaffold gaps somewhat tolerant of different alignment scenarios.

For any P a *scaffold path* is a maximal subpath of P in which all the edges are consistent and/or part of a scaffold gap subgraph. As a scaffold path is simply a concatenation of haplotype paths its base pair length is simply defined as the sum of the base pair lengths of the haplotype paths that it contains.

2.3.4 Errors

For any thread P representing an assembly sequence an edge e in P is called *erroneous* if it is not part of a scaffold path (by definition any edge is a member of at most one scaffold path). To categorise these errors we have defined a number of subgraphs, each involving a path of erroneous edges, these are illustrated in Figure 2.7.

2.3.5 NA50

The *N50* of an assembly is a weighted median equal to the size of a sequence s in the assembly such that the sum of the lengths of sequences in the assembly equal or greater in length than s cover half the genome being assembled. As the length of the genome being assembled is normally unknown, or, as in this case, difficulty to establish given the polymorphism of the two haplotypes, we use as a proxy the total length of the sequences in the assembly.

We define the *NA50* (A for alignment) identically to the N50, except that we estimate the length of the genome being assembled as being equal to the total number of columns in the MSA containing positions in the haplotype sequences. The *SPA50*, *HPA50* and *BA50* values are identical to the NA50s, except that they are computed over the set of scaffold paths, haplotype paths and blocks, respectively.

2.3.6 Substitution errors

Although we do not allow structural rearrangements within blocks, blocks are tolerant of substitutions. Let a column of aligned bases within a block that (1) contains

a single position from one or both haplotypes and (2) a single position from an assembly sequence be called *valid*. We use this criteria because such columns unambiguously map a single assembled sequence to a single position in the alignment of the two haplotypes, while avoiding the issues of paralogous alignment and multiple counting. We distinguish three types of valid columns: (1) *Homozygous columns*, those containing a member of both haplotypes, both of which have the same basepair, (2) *Heterozygous columns*, those containing a member of both haplotypes, but which have distinct basepairs, and (3) *Indels columns*, those containing a member of either haplotype, but not both.

Assemblers are free to use IUPAC ambiguity characters to call bases. We use a bit-score to score correct but ambiguous matches within valid columns. We assign the column a score of $m - \log_2(n/4)$, where n is the number of different bases the IUPAC character in the assembly represents and m is the number of distinct base pairs in the two haplotypes that matches or is represented (amongst others) by the assembly IUPAC character. Thus in homozygous and indel columns the score is at most 2, in heterozygous columns the score is 2 if and only if the assembly correctly predicts one of the two base pairs, or if it predicts an ambiguity character which represents both and only those two base pairs. We say there has been a *substitution error* if the assembly sequence's position has a IUPAC character that is not one of the haplotype position's base pair(s).

Some of the substitution errors that we observe are like due to misalignments. These can occur due to edge wander, or the larger scale misalignment of an assembled sequence to a paralog of its true ortholog. The sum of substitution errors over all valid columns is therefore an upper bound on the substitution errors within valid columns. To obtain a higher confidence set of substitution errors we select a subset of valid columns that meet the following requirements. (1) Are part of blocks of at least 100 base pairs in length, avoiding errors within short indels, (2) are not within 5 positions of the start and end of the block, avoiding edge wander, and (3) are within sequences of 98% or higher identity, ensuring the alignments are unlikely to be paralogous. The sum of substitution errors within these high confidence valid columns represents a reasonable lower bound of the number of substitution errors.

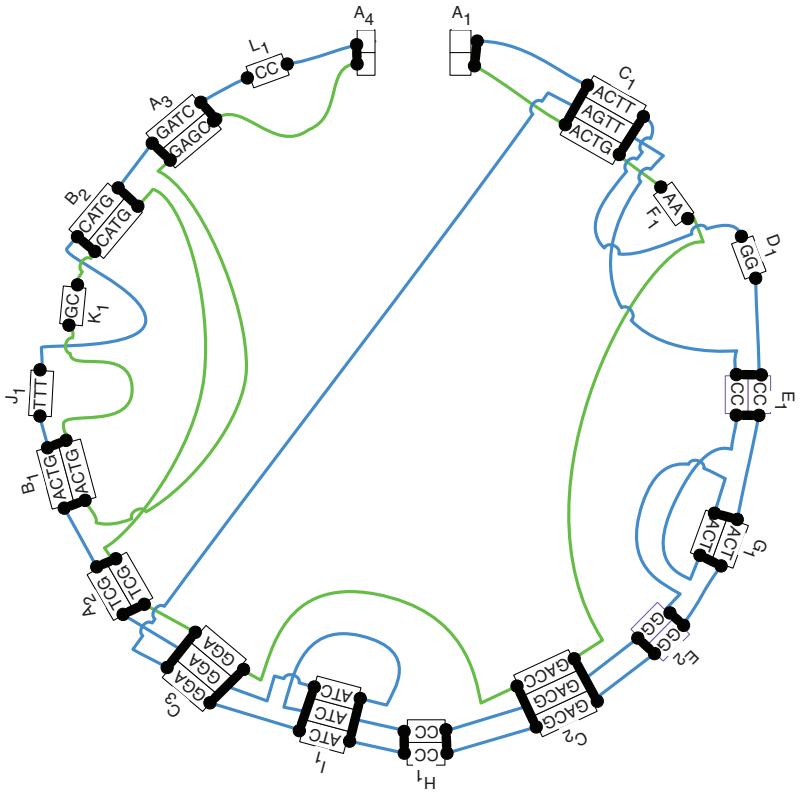


Figure 2.4: A circular genome style plot showing an adjacency graph, with examples of contained threads. Blue and green lines depict two homologous threads traversing a series of segments in blocks and the joining adjacencies. All aligned boxes represent block edges except A_1 and A_4, which are ends representing telomeres. The ends of the blocks and the ends of the telomeres are mapped as filled black rectangles on the edges of the aligned boxes. These are the nodes in the adjacency graph. The adjacency edges are represented as collections of colored lines connecting the ends of blocks. Starting at A_1, the blue thread gives the sequence ACTTGGCCACTGGGACGCCATCGGAAGTTCCagtGGGACGCCATCATCGGATCGACTGTTTCATGGATCCC. The green thread gives the sequence ACTGAAGACCGGATCGcatggccagtGAGC. The lower case "agt" in the blue thread represents the reverse complement of the bottom segment of block G_1, which is traversed right-to-left in the blue thread and similarly, the lower case segment in the green thread is the reverse complement of segments in B_2, K_1 and B_1, also traversed right-to-left. This figure is adapted from [Paten et al., 2011]

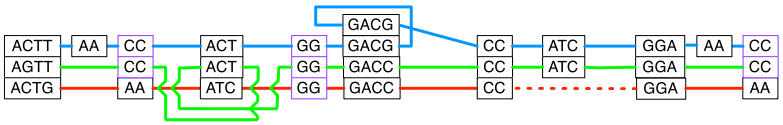


Figure 2.5: A subgraph of an adjacency graph, represented as in Figure 2.4. The blue and green lines depict two homologous threads of the two haplotypes, the red lines represents the thread of an assembly sequence. The red thread is composed of two haplotype paths linked by the dotted red line, which represents an error in the assembly as it is not consistent with either haplotype.

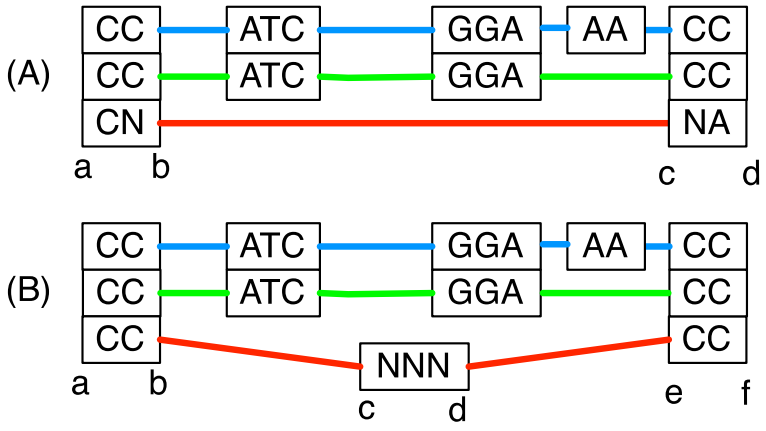


Figure 2.6: An illustration of the two types of scaffold gap within subgraphs of an adjacency graph. The blue and green lines depict two homologous threads of the two haplotypes, the red lines represents the thread of an assembly sequence. (a) A scaffold gap without unaligned sequence. The nodes of the scaffold gap are labelled as described in the main text. (b) A scaffold gap without aligned sequence, again the nodes of the scaffold gap are labelled as described in the main text.

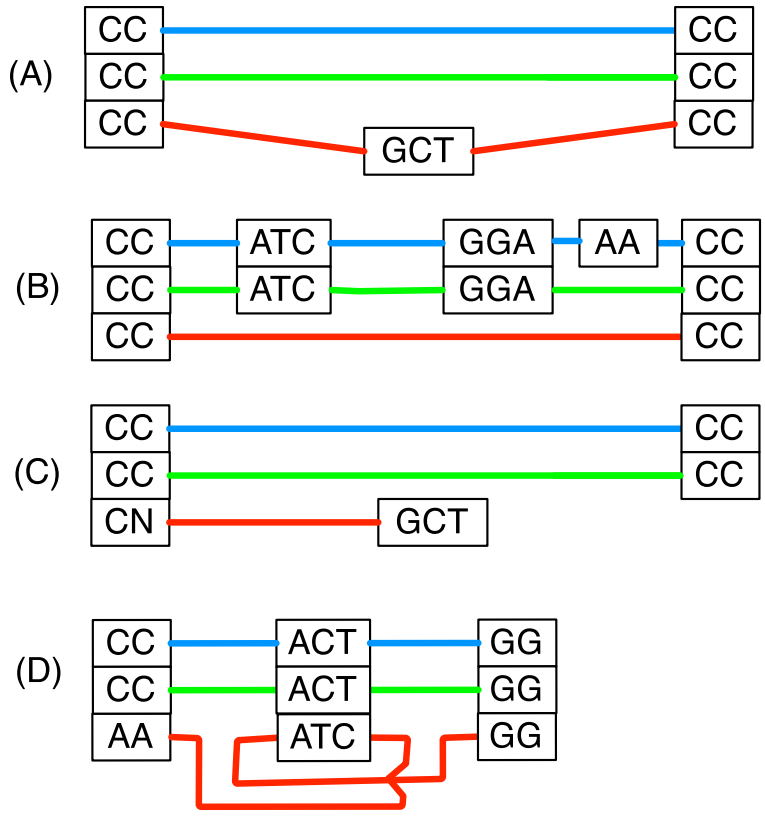


Figure 2.7: An illustration of different error structures recognisable in the adjacency graph. (a) An insertion error. (B) A deletion error. (C) A hanging end error. (D) A haplotype joining error. Each such structure can be recognised as a characteristic subgraph of an adjacency graph.

Chapter 3

Results

3.1 Global results

Table 3.1 *(following page)*: Assembly coverage. Total coverage is calculated as the number of columns aligned to haplotype 1 plus the number of columns aligned to haplotype 2 divided by the sum of the columns in haplotypes 1 and 2. Delta is absolute difference between the Hap 1 and Hap 2 columns. The Bac column is the assemblies coverage against the genome of the bacterial contamination.

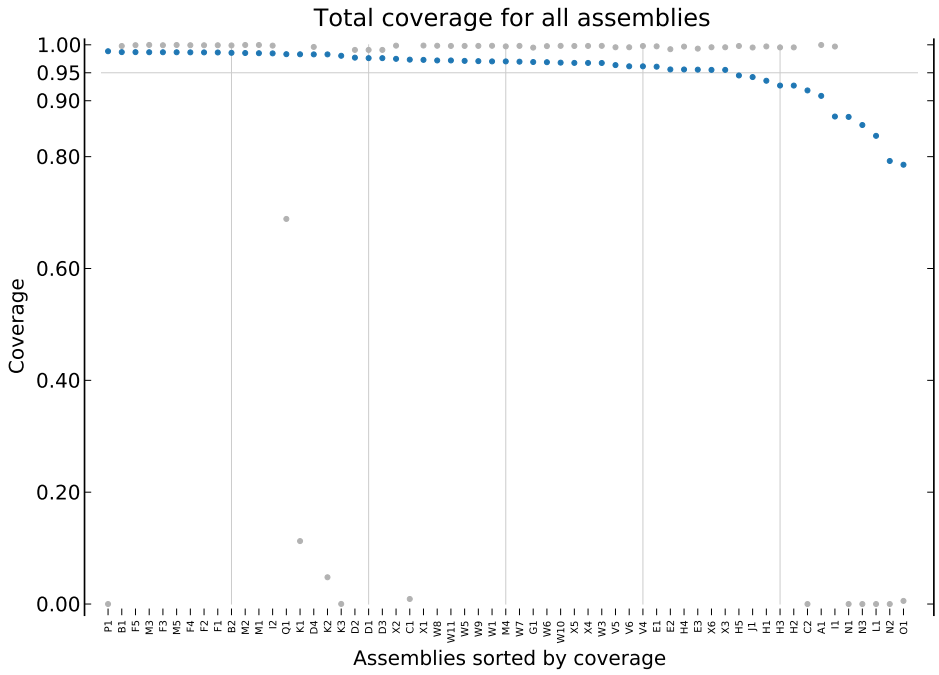


Figure 3.1: Total coverage, as calculated in Table 3.1, is shown in blue and bacterial contaimination coverage is shown in grey.

Assembly	Total	Hap 1	Hap 2	Delta	Bac
P1	0.9885	0.9888	0.9882	5.782e-04	0
B1	0.9869	0.9871	0.9866	5.257e-04	0.9978
F5	0.9869	0.9872	0.9865	7.295e-04	0.9993
M3	0.9867	0.9868	0.9866	2.428e-04	0.9997
F3	0.9867	0.9869	0.9864	4.757e-04	0.9992
M5	0.9866	0.9867	0.9865	2.526e-04	0.9996
F4	0.9864	0.9867	0.9862	5.014e-04	0.9992
F2	0.9864	0.9867	0.9861	6.557e-04	0.9992
F1	0.9862	0.9866	0.9859	6.860e-04	0.9992
B2	0.9856	0.9859	0.9853	6.450e-04	0.9989
M2	0.9853	0.9855	0.9851	3.559e-04	0.9996
M1	0.9849	0.9852	0.9847	4.572e-04	0.9996
I2	0.9846	0.9851	0.9842	8.726e-04	0.9985
Q1	0.9832	0.9833	0.9831	2.598e-04	0.6886
K1	0.9830	0.9830	0.9830	6.381e-05	0.1125
D4	0.9828	0.9830	0.9827	2.812e-04	0.9961
K2	0.9828	0.9828	0.9828	6.380e-06	0.0478
K3	0.9803	0.9804	0.9803	9.198e-05	0.0001
D2	0.9770	0.9771	0.9769	1.960e-04	0.9908
D1	0.9761	0.9764	0.9759	5.247e-04	0.9908
D3	0.9761	0.9763	0.9759	3.876e-04	0.9908
X2	0.9749	0.9751	0.9746	4.908e-04	0.9984
C1	0.9734	0.9734	0.9734	1.242e-05	0.0090
X1	0.9730	0.9733	0.9727	5.584e-04	0.9986
W8	0.9720	0.9723	0.9717	5.625e-04	0.9983
W11	0.9720	0.9722	0.9718	3.291e-04	0.9979
W5	0.9712	0.9715	0.9710	5.088e-04	0.9980
W9	0.9707	0.9710	0.9705	4.360e-04	0.9980
W1	0.9703	0.9704	0.9702	2.492e-04	0.9982
M4	0.9703	0.9704	0.9701	3.194e-04	0.9971
W7	0.9698	0.9700	0.9696	4.411e-04	0.9980
G1	0.9692	0.9695	0.9689	5.803e-04	0.9949
W6	0.9689	0.9691	0.9686	5.379e-04	0.9976
W10	0.9681	0.9685	0.9677	7.935e-04	0.9981
X5	0.9676	0.9678	0.9674	4.724e-04	0.9978
X4	0.9675	0.9677	0.9673	4.354e-04	0.9979
W3	0.9675	0.9677	0.9672	4.277e-04	0.9981
V5	0.9637	0.9638	0.9635	3.257e-04	0.9955
V6	0.9615	0.9618	0.9613	5.148e-04	0.9955
V4	0.9615	0.9617	0.9612	5.129e-04	0.9978
E1	0.9608	0.9612	0.9605	7.138e-04	0.9972
E2	0.9559	0.9563	0.9555	8.815e-04	0.9921
H4	0.9558	0.9558	0.9558	3.405e-05	0.9968
E3	0.9555	0.9560	0.9551	8.369e-04	0.9931
X6	0.9551	0.9552	0.9550	2.358e-04	0.9956
X3	0.9551	0.9553	0.9549	4.028e-04	0.9955
H5	0.9451	0.9453	0.9450	2.885e-04	0.9978
J1	0.9423	0.9424	0.9421	3.011e-04	0.9951
H1	0.9357	0.9358	0.9356	1.461e-04	0.9970
H3	0.9271	0.9273	0.9269	3.544e-04	0.9953
H2	0.9271	0.9272	0.9269	3.341e-04	0.9953
C2	0.9184	0.9187	0.9180	7.282e-04	0
A1	0.9086	0.9087	0.9085	2.653e-04	0.9997
I1	0.8717	0.8721	0.8713	7.594e-04	0.9969
N1	0.8710	0.8712	0.8709	2.700e-04	0
N3	0.8565	0.8568	0.8563	4.563e-04	0
L1	0.8372	0.8372	0.8371	1.148e-04	0
N2	0.7922	0.7925	0.7919	6.170e-04	0
O1	0.7855	0.7901	0.7809	9.206e-03	0.0055

Figure 3.2 *(following page)*: Bacterial contamination gapless block length coverages. The annotation shown is he union of Tandem Repeats Finder and RepeatMasker output (repeat). Red lines at the top of the tracks indicate the track goes off the scale and is visually clipped.
Assemblies are sorted in order of percent coverage, caluclated as the number of columns aligning to the haplotype divided by the length of the haplotype.
The color of the curve is indicative of the proportion of bases in that area of the genome that are in gapless blocks of a certain size threshold. The colored thresholds key is printed at the bottom. Feature positions have been discretized into 2,400 bins ($\sim$ 47 Kb/bin) for display.

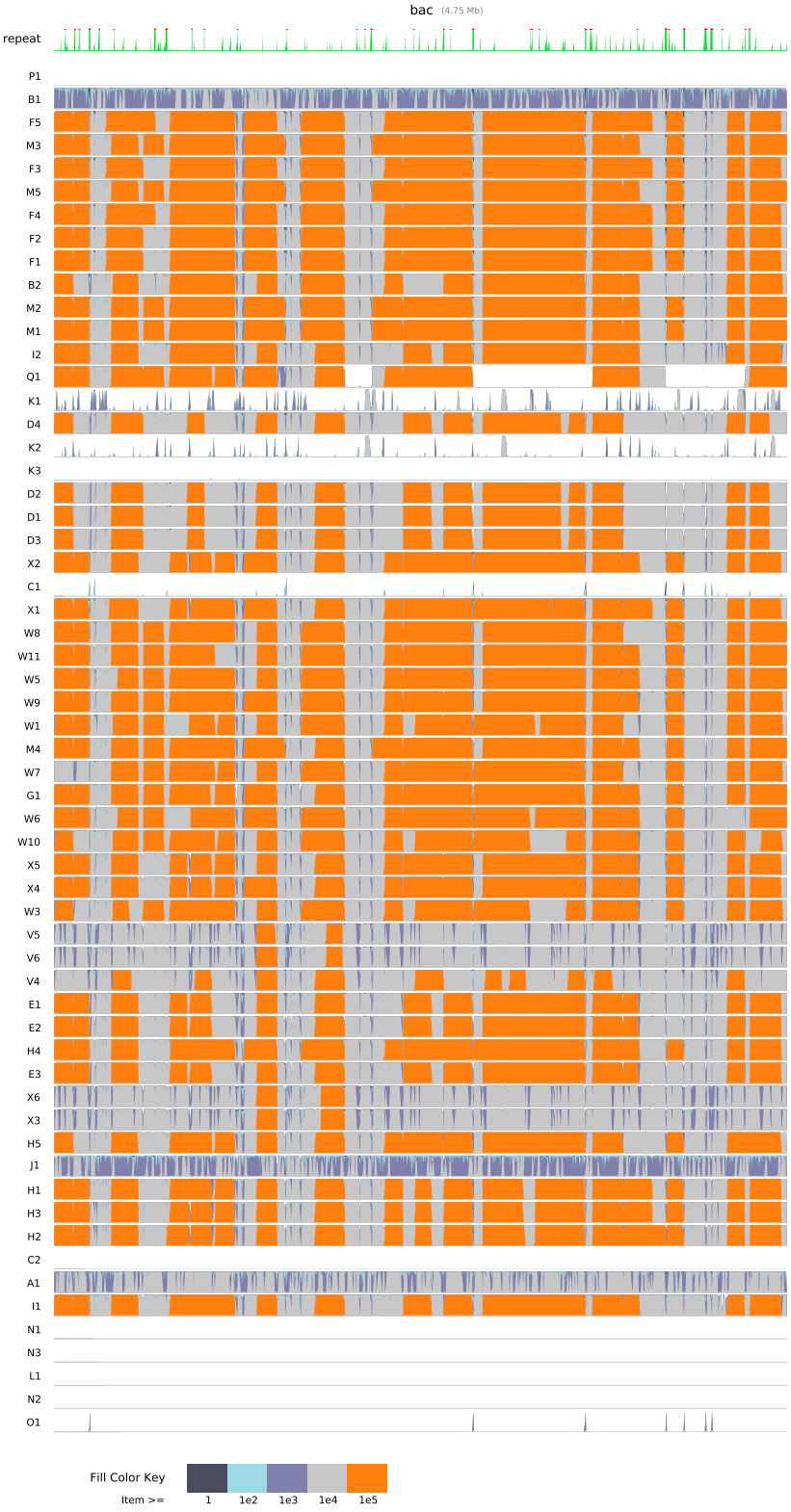

bac (4.75 Mb)
repeat
P1
B1
F5
M3
F3
M5
F4
F2
F1
B2
M2
M1
I2
Q1
K1
D4
K2
K3
D2
D1
D3
X2
C1
X1
W8
W11
W5
W9
W1
M4
W7
G1
W6
W10
X5
X4
W3
V5
V6
V4
E1
E2
H4
E3
X6
X3
H5
J1
H1
H3
H2
C2
A1
I1
N1
N3
L1
N2
O1
Fill Color Key
Item >=
1
1e2
1e3
1e4
1e5

Figure 3.3 *(following page)*: Haplotype 1 contig length coverages. Annotations shown are (top-bottom) coding sequence (CDS), untranslated region (UTR), non exonic constrained elements (NXE), non genic constrained elements (NGE), CpG islands (island) and the union of Tandem Repeats Finder and RepeatMasker output (repeat). Red lines at the top of the tracks indicate the track goes off the scale and is visually clipped.
Assemblies are sorted in order of percent coverage, caluclated as the number of columns aligning to the haplotype divided by the length of the haplotype.
The color of the curve is indicative of the proportion of bases in that area of the genome that are in contigs of a certain size threshold. The colored thresholds key is printed at the bottom. Feature positions have been discretized into 2,400 bins ($\sim$ 47 Kb/bin) for display.

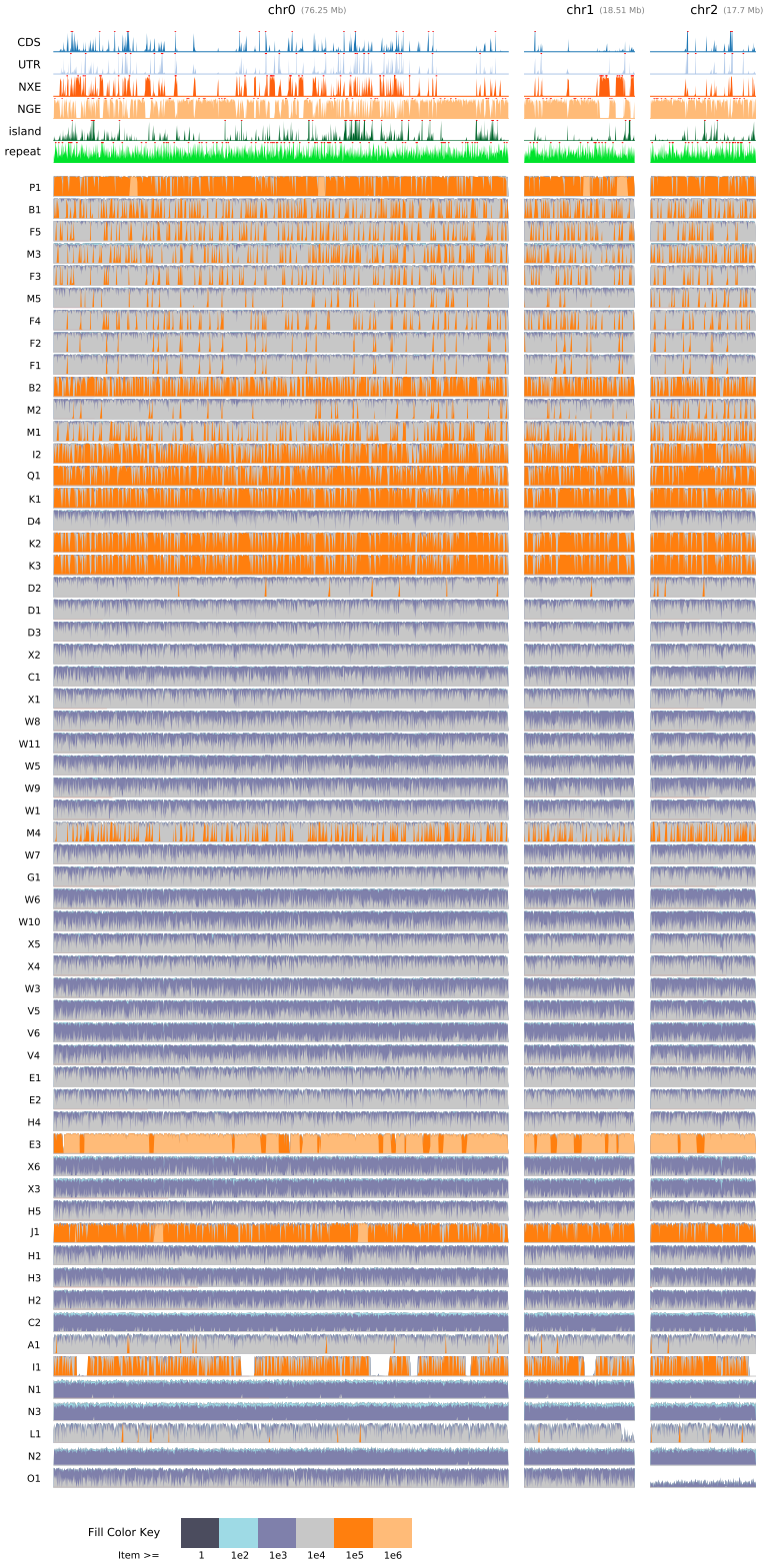

chr0 (76.25 Mb)
chr1 (18.51 Mb)
chr2 (17.7 Mb)
CDS
UTR
NXE
NGE
island
repeat
P1
B1
F5
M3
F3
M5
F4
F2
F1
B2
M2
M1
I2
Q1
K1
D4
K2
K3
D2
D1
D3
X2
C1
X1
W8
W11
W5
W9
W1
M4
W7
G1
W6
W10
X5
X4
W3
V5
V6
V4
E1
E2
H4
E3
X6
X3
H5
J1
H1
H3
H2
C2
A1
I1
N1
N3
L1
N2
O1
Fill Color Key
Item >=
1
1e2
1e3
1e4
1e5
1e6

Figure 3.4 *(following page)*: Haplotype 2 contig length coverages. Annotations shown are (top-bottom) coding sequence (CDS), untranslated region (UTR), non exonic constrained elements (NXE), non genic constrained elements (NGE), CpG islands (island) and the union of Tandem Repeats Finder and RepeatMasker output (repeat). Red lines at the top of the tracks indicate the track goes off the scale and is visually clipped.
Assemblies are sorted in order of percent coverage, caluclated as the number of columns aligning to the haplotype divided by the length of the haplotype.
The color of the curve is indicative of the proportion of bases in that area of the genome that are in contigs of a certain size threshold. The colored thresholds key is printed at the bottom. Feature positions have been discretized into 2,400 bins ($\sim$ 47 Kb/bin) for display.

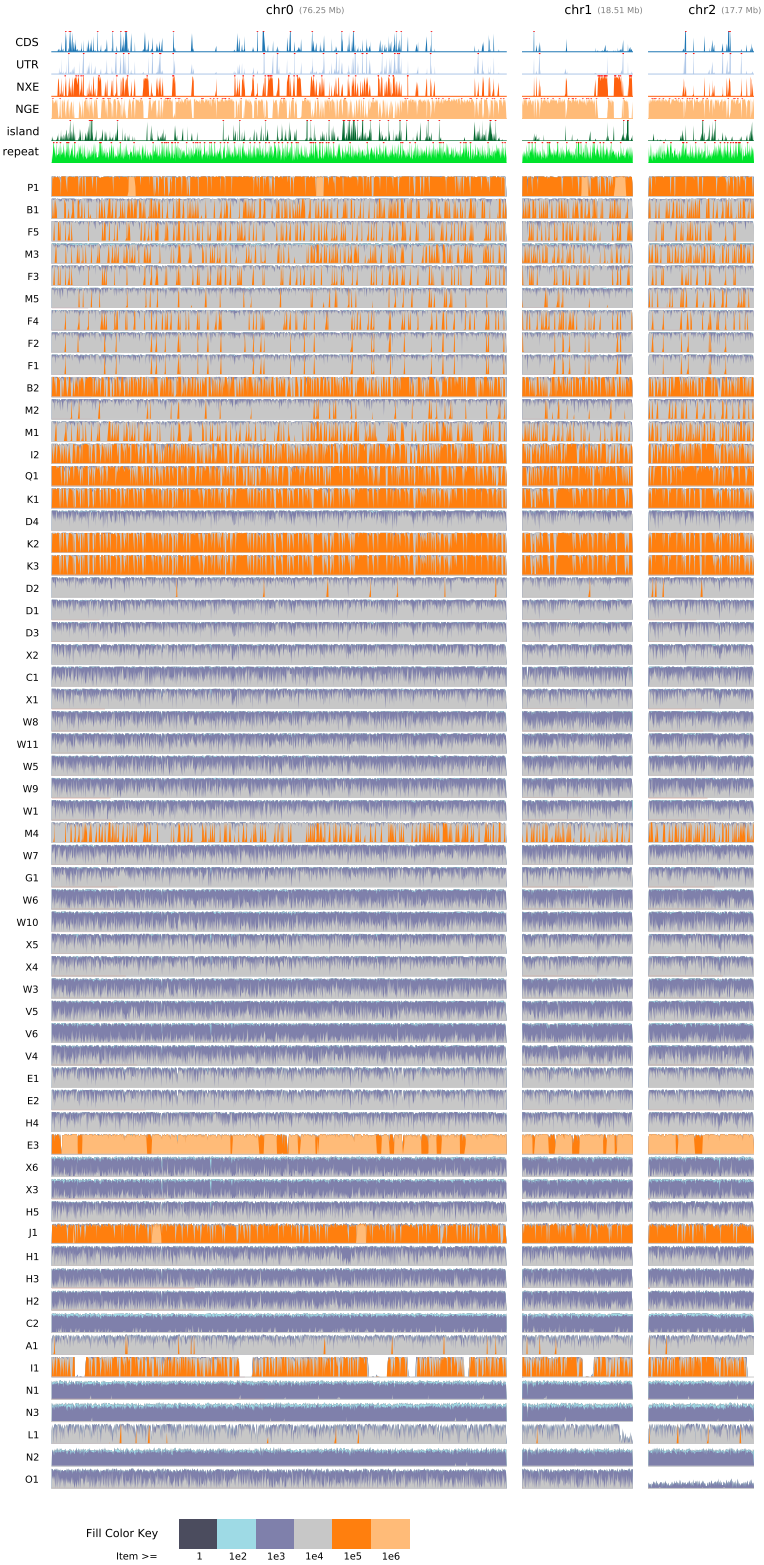

chr0 (76.25 Mb)
chr1 (18.51 Mb)
chr2 (17.7 Mb)
CDS
UTR
NXE
NGE
island
repeat
P1
B1
F5
M3
F3
M5
F4
F2
F1
B2
M2
M1
I2
Q1
K1
D4
K2
K3
D2
D1
D3
X2
C1
X1
W8
W11
W5
W9
W1
M4
W7
G1
W6
W10
X5
X4
W3
V5
V6
V4
E1
E2
H4
E3
X6
X3
H5
J1
H1
H3
H2
C2
A1
I1
N1
N3
L1
N2
O1
Fill Color Key
Item >=
1
1e2
1e3
1e4
1e5
1e6

Figure 3.5 *(following page)*: Haplotype 1 scaffold path length coverages. Annotations shown are (top-bottom) coding sequence (CDS), untranslated region (UTR), non exonic constrained elements (NXE), non genic constrained elements (NGE), CpG islands (island) and the union of Tandem Repeats Finder and RepeatMasker output (repeat). Red lines at the top of the tracks indicate the track goes off the scale and is visually clipped.
Assemblies are sorted in order of percent coverage, caluclated as the number of columns aligning to the haplotype divided by the length of the haplotype.
The color of the curve is indicative of the proportion of bases in that area of the genome that are in scaffold paths of a certain size threshold. The colored thresholds key is printed at the bottom. Feature positions have been discretized into 2,400 bins ($\sim$ 47 Kb/bin) for display.

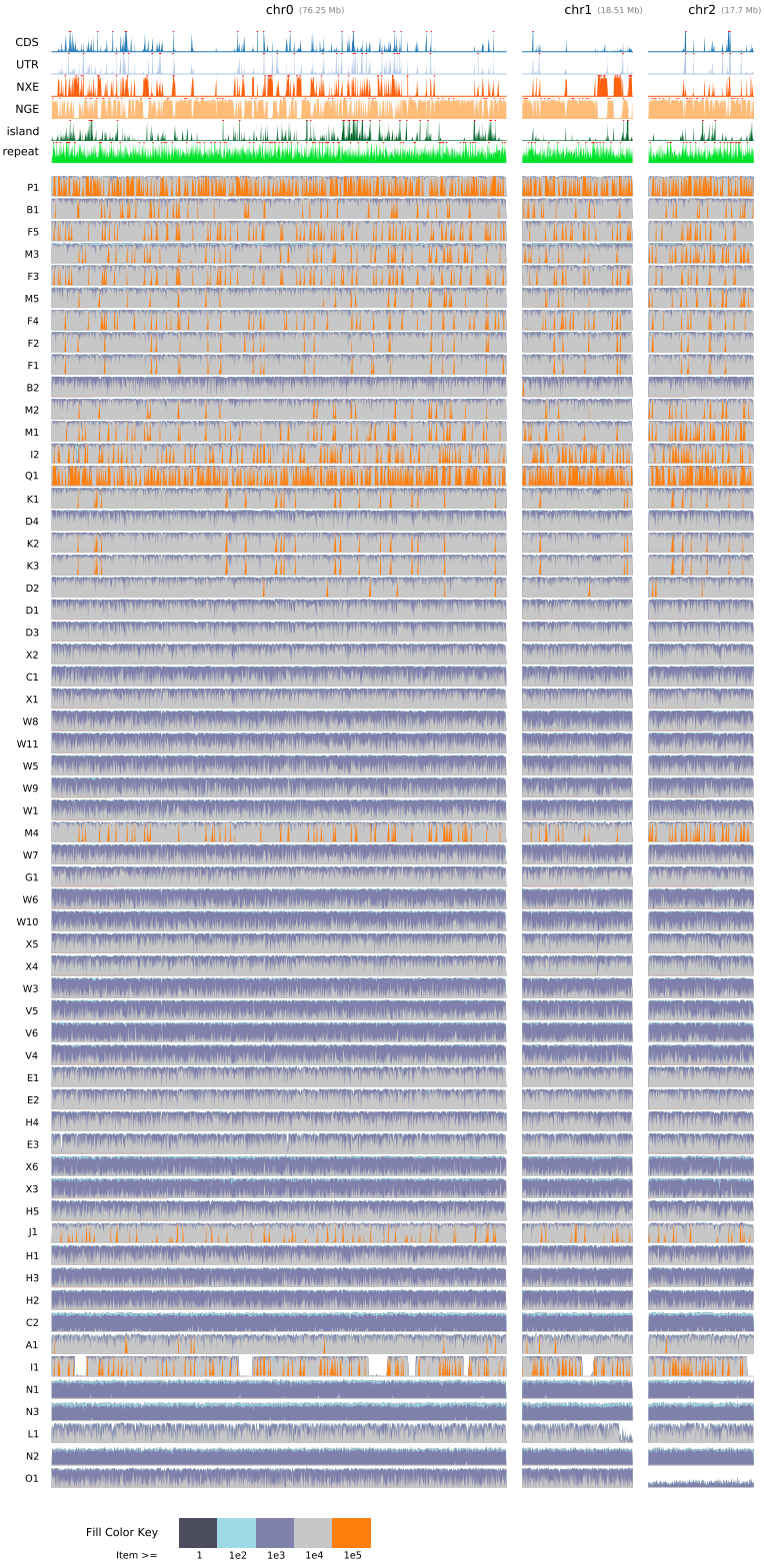

chr0 (76.25 Mb)
chr1 (18.51 Mb)
chr2 (17.7 Mb)
CDS
UTR
NXE
NGE
island
repeat
P1
B1
F5
M3
F3
M5
F4
F2
F1
B2
M2
M1
I2
Q1
K1
D4
K2
K3
D2
D1
D3
X2
C1
X1
W8
W11
W5
W9
W1
M4
W7
G1
W6
W10
X5
X4
W3
V5
V6
V4
E1
E2
H4
E3
X6
X3
H5
J1
H1
H3
H2
C2
A1
I1
N1
N3
L1
N2
O1
Fill Color Key
Item >=
1
1e2
1e3
1e4
1e5

Figure 3.6 *(following page)*: Haplotype 2 scaffold path length coverages. Annotations shown are (top-bottom) coding sequence (CDS), untranslated region (UTR), non exonic constrained elements (NXE), non genic constrained elements (NGE), CpG islands (island) and the union of Tandem Repeats Finder and RepeatMasker output (repeat). Red lines at the top of the tracks indicate the track goes off the scale and is visually clipped.
Assemblies are sorted in order of percent coverage, caluclated as the number of columns aligning to the haplotype divided by the length of the haplotype.
The color of the curve is indicative of the proportion of bases in that area of the genome that are in scaffold paths of a certain size threshold. The colored thresholds key is printed at the bottom. Feature positions have been discretized into 2,400 bins ($\sim$ 47 Kb/bin) for display.

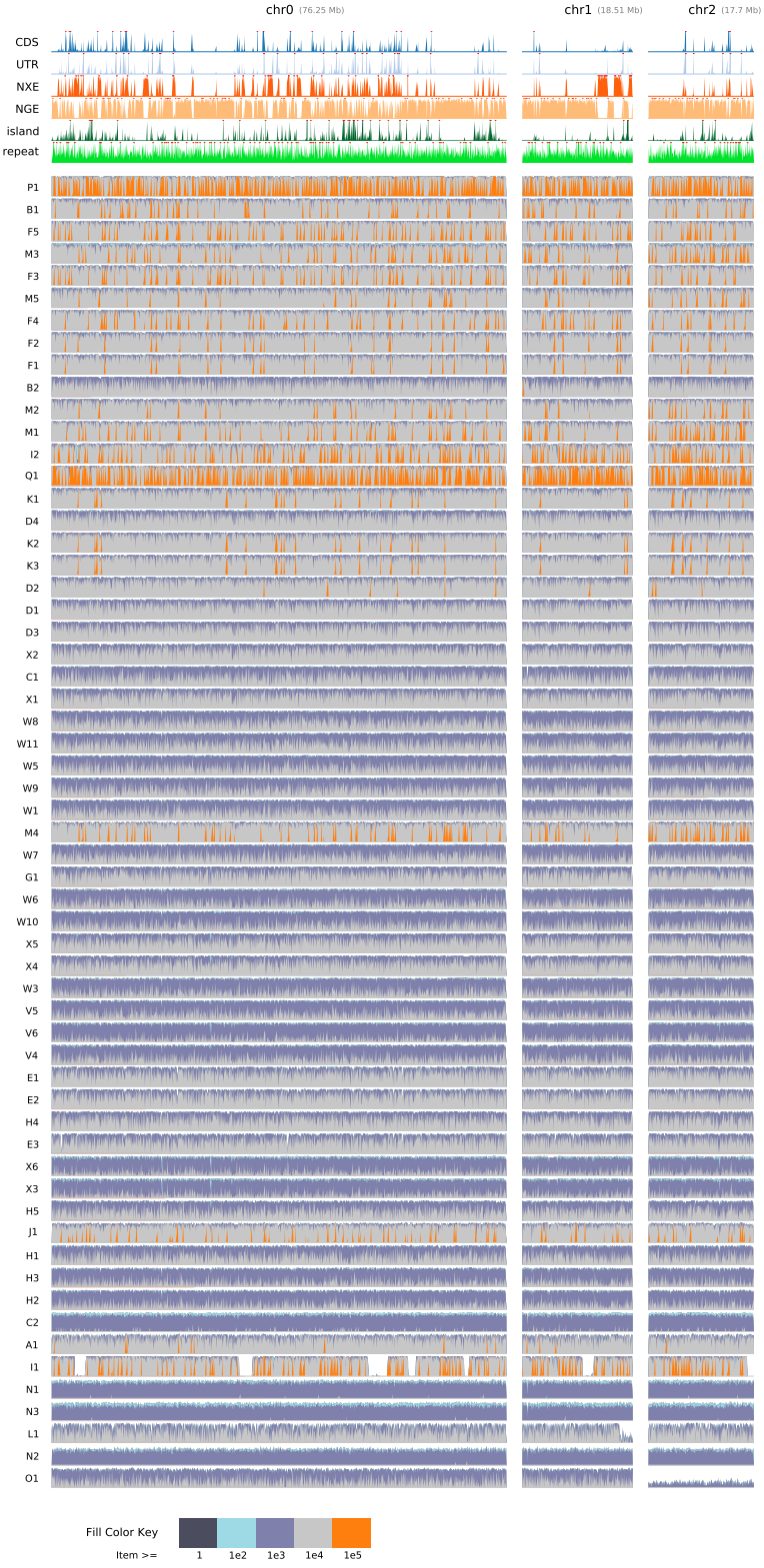

chr0 (76.25 Mb)
chr1 (18.51 Mb)
chr2 (17.7 Mb)
CDS
UTR
NXE
NGE
island
repeat
P1
B1
F5
M3
F3
M5
F4
F2
F1
B2
M2
M1
I2
Q1
K1
D4
K2
K3
D2
D1
D3
X2
C1
X1
W8
W11
W5
W9
W1
M4
W7
G1
W6
W10
X5
X4
W3
V5
V6
V4
E1
E2
H4
E3
X6
X3
H5
J1
H1
H3
H2
C2
A1
I1
N1
N3
L1
N2
O1
Fill Color Key
Item >=
1
1e2
1e3
1e4
1e5

Figure 3.7 *(following page)*: Haplotype 1 haplotype path length coverages. Annotations shown are (top-bottom) coding sequence (CDS), untranslated region (UTR), non exonic constrained elements (NXE), non genic constrained elements (NGE), CpG islands (island) and the union of Tandem Repeats Finder and RepeatMasker output (repeat). Red lines at the top of the tracks indicate the track goes off the scale and is visually clipped.
Assemblies are sorted in order of percent coverage, caluclated as the number of columns aligning to the haplotype divided by the length of the haplotype.
The color of the curve is indicative of the proportion of bases in that area of the genome that are in haplotype paths of a certain size threshold. The colored thresholds key is printed at the bottom. Feature positions have been discretized into 2,400 bins ($\sim$ 47 Kb/bin) for display.

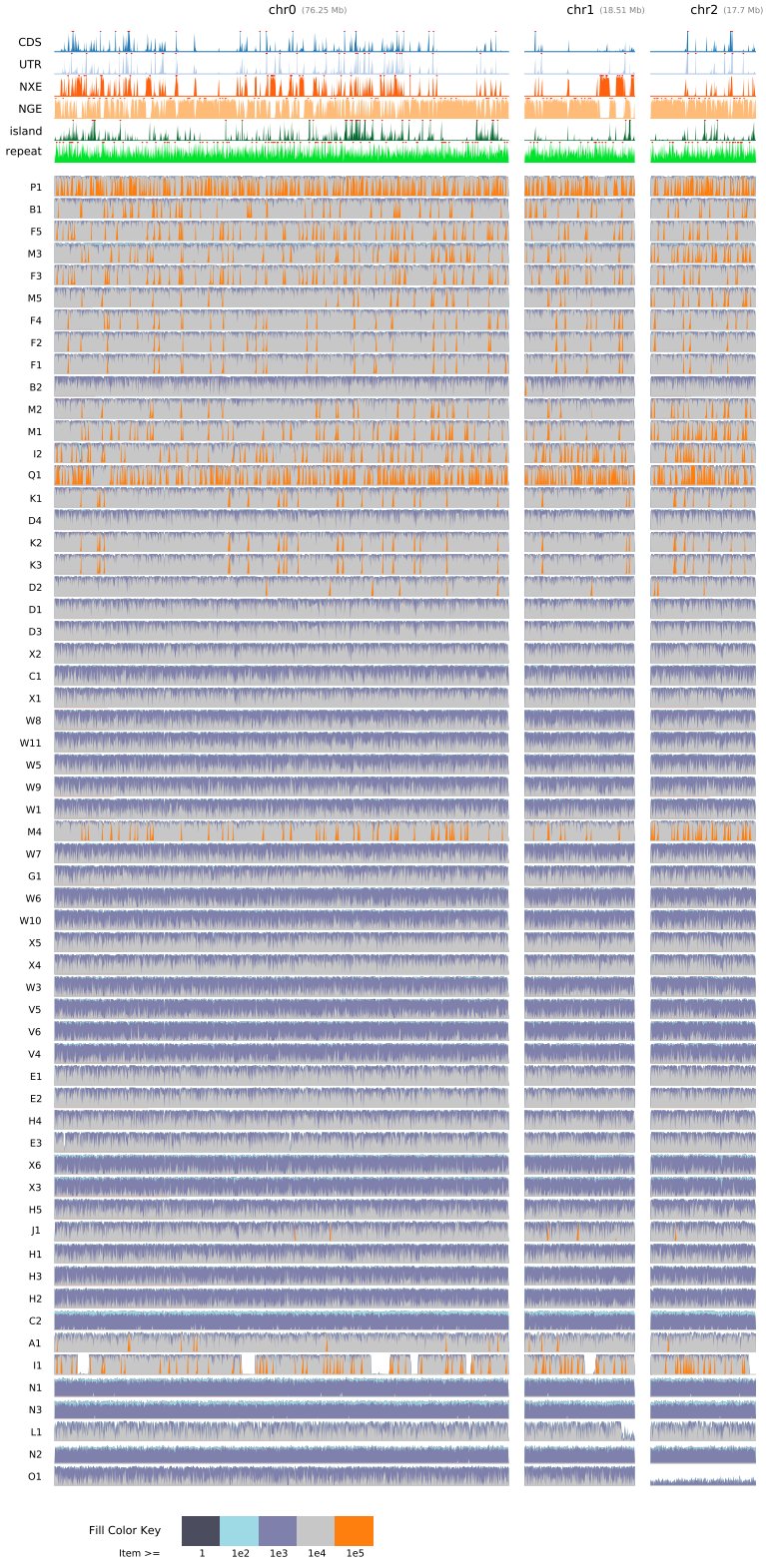

chr0 (76.25 Mb)
chr1 (18.51 Mb)
chr2 (17.7 Mb)
CDS
UTR
NXE
NGE
island
repeat
P1
B1
F5
M3
F3
M5
F4
F2
F1
B2
M2
M1
I2
Q1
K1
D4
K2
K3
D2
D1
D3
X2
C1
X1
W8
W11
W5
W9
W1
M4
W7
G1
W6
W10
X5
X4
W3
V5
V6
V4
E1
E2
H4
E3
X6
X3
H5
J1
H1
H3
H2
C2
A1
I1
N1
N3
L1
N2
O1
Fill Color Key
Item >=
1
1e2
1e3
1e4
1e5

Figure 3.8 *(following page)*: Haplotype 2 haplotype path length coverages. Annotations shown are (top-bottom) coding sequence (CDS), untranslated region (UTR), non exonic constrained elements (NXE), non genic constrained elements (NGE), CpG islands (island) and the union of Tandem Repeats Finder and RepeatMasker output (repeat). Red lines at the top of the tracks indicate the track goes off the scale and is visually clipped.
Assemblies are sorted in order of percent coverage, caluclated as the number of columns aligning to the haplotype divided by the length of the haplotype.
The color of the curve is indicative of the proportion of bases in that area of the genome that are in haplotype paths of a certain size threshold. The colored thresholds key is printed at the bottom. Feature positions have been discretized into 2,400 bins ($\sim$ 47 Kb/bin) for display.

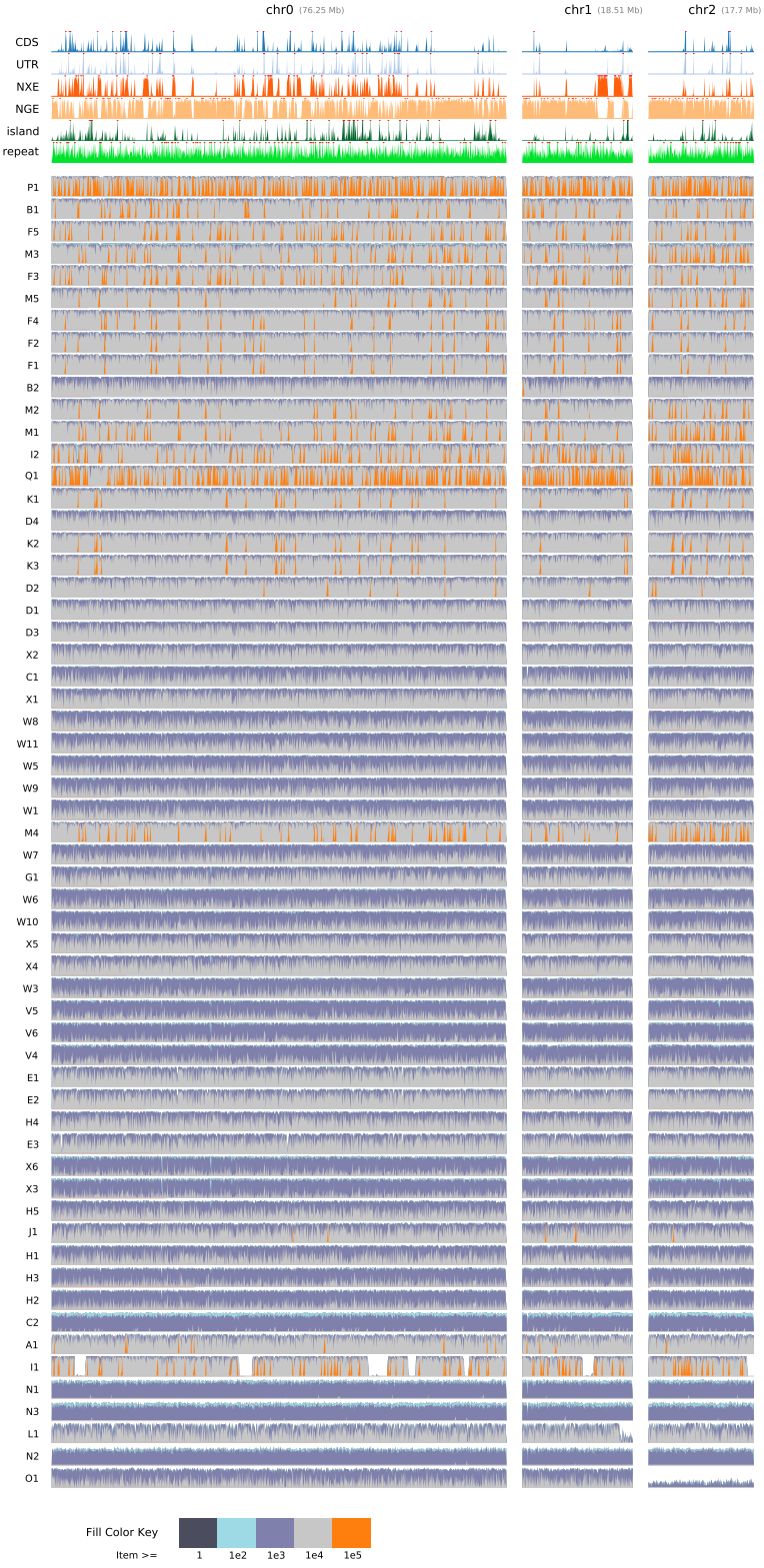

chr0 (76.25 Mb)
chr1 (18.51 Mb)
chr2 (17.7 Mb)
CDS
UTR
NXE
NGE
island
repeat
P1
B1
F5
M3
F3
M5
F4
F2
F1
B2
M2
M1
I2
Q1
K1
D4
K2
K3
D2
D1
D3
X2
C1
X1
W8
W11
W5
W9
W1
M4
W7
G1
W6
W10
X5
X4
W3
V5
V6
V4
E1
E2
H4
E3
X6
X3
H5
J1
H1
H3
H2
C2
A1
I1
N1
N3
L1
N2
O1
Fill Color Key
Item >=
1
1e2
1e3
1e4
1e5

Figure 3.9 *(following page)*: Haplotype 1 annotations and assemblies sorted on coverage. Annotations shown are (top-bottom) coding sequence (CDS), untranslated region (UTR), non exonic constrained elements (NXE), non genic constrained elements (NGE), CpG islands (island) and the union of Tandem Repeats Finder and RepeatMasker output (repeat). Red lines at the top of the tracks indicate the track goes off the scale and is visually clipped.
Assemblies are sorted in order of percent coverage, calculated as the number of columns aligning to the haplotype divided by the length of the haplotype.
The black line represents local coverage. The blue line represents haplotype edge density. The red line represents haplotype error density. Every assembly shows the haplotype edge and error densities on the same scale to permit between assembly comparison by eye. Feature positions have been discretized into 2,400 bins ($\sim$ 47 Kb/bin) for display.

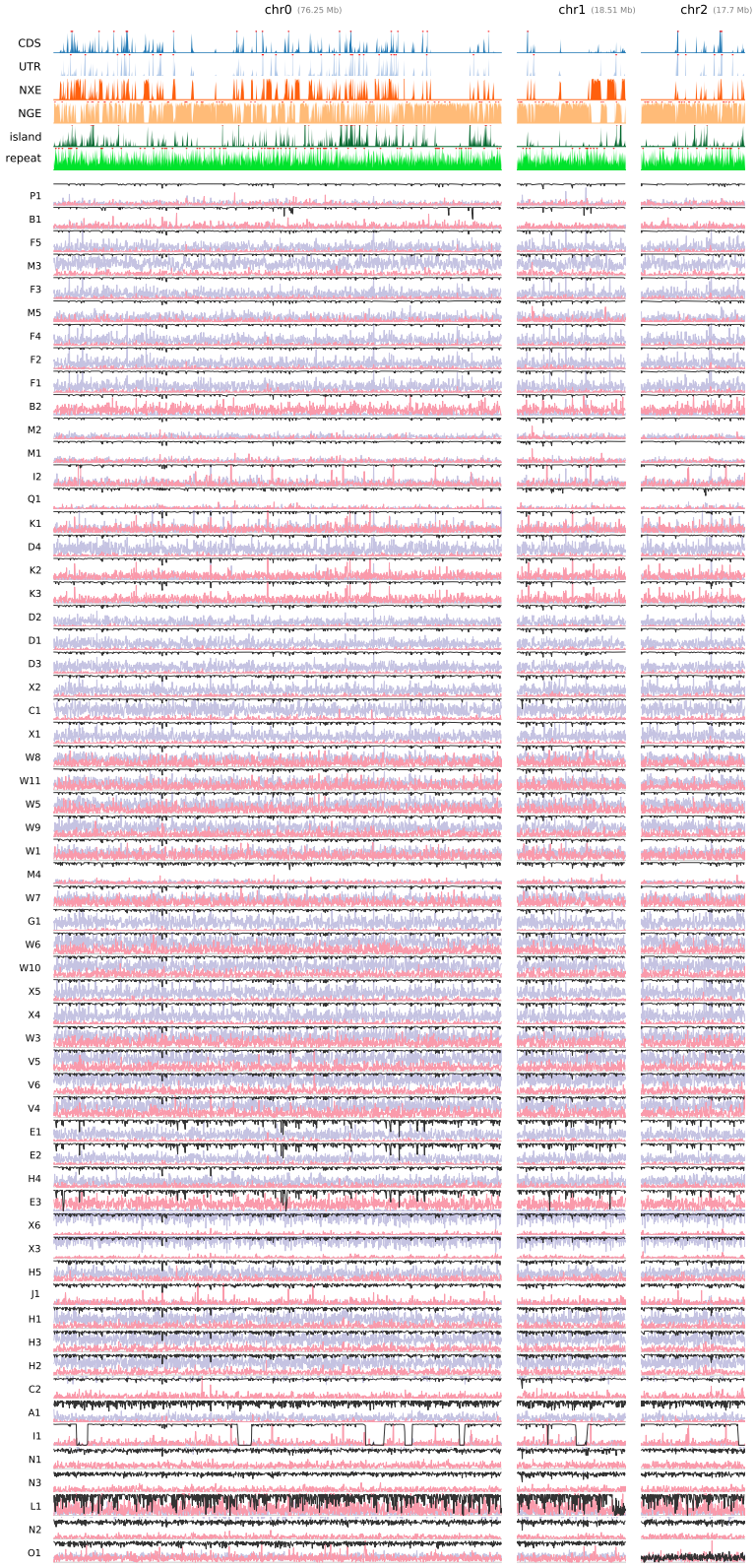

chr0 (76.25 Mb)
chr1 (18.51 Mb)
chr2 (17.7 Mb)
CDS
UTR
NXE
NGE
island
repeat
P1
B1
F5
M3
F3
M5
F4
F2
F1
B2
M2
M1
I2
Q1
K1
D4
K2
K3
D2
D1
D3
X2
C1
X1
W8
W11
W5
W9
W1
M4
W7
G1
W6
W10
X5
X4
W3
V5
V6
V4
E1
E2
H4
E3
X6
X3
H5
J1
H1
H3
H2
C2
A1
I1
N1
N3
L1
N2
O1

Figure 3.10 *(following page)*: Haplotype 2 annotations and assemblies sorted on coverage. Annotations shown are (top-bottom) coding sequence (CDS), untranslated region (UTR), non exonic constrained elements (NXE), non genic constrained elements (NGE), CpG islands (island) and the union of Tandem Repeats Finder and RepeatMasker output (repeat). Red lines at the top of the tracks indicate the track goes off the scale and is visually clipped.
Assemblies are sorted in order of percent coverage, calculated as the number of columns aligning to the haplotype divided by the length of the haplotype.
The black line represents local coverage. The blue line represents haplotype edge density. The red line represents haplotype error density. Every assembly shows the haplotype edge and error densities on the same scale to permit between assembly comparison by eye. Feature positions have been discretized into 2,400 bins ($\sim$ 47 Kb/bin) for display.

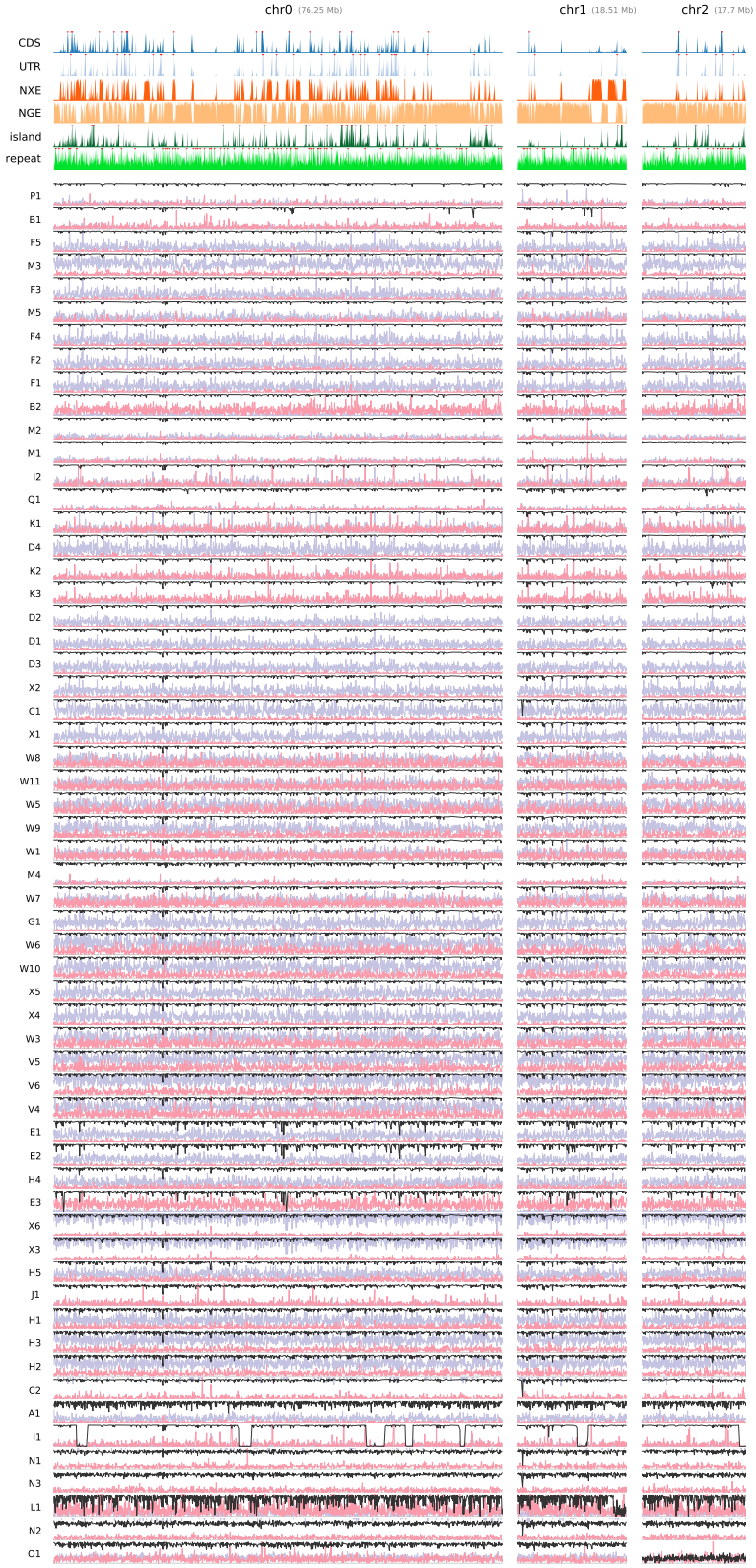

chr0 (76.25 Mb)
chr1 (18.51 Mb)
chr2 (17.7 Mb)
CDS
UTR
NXE
NGE
island
repeat
P1
B1
F5
M3
F3
M5
F4
F2
F1
B2
M2
M1
I2
Q1
K1
D4
K2
K3
D2
D1
D3
X2
C1
X1
W8
W11
W5
W9
W1
M4
W7
G1
W6
W10
X5
X4
W3
V5
V6
V4
E1
E2
H4
E3
X6
X3
H5
J1
H1
H3
H2
C2
A1
I1
N1
N3
L1
N2
O1

Figure 3.11 *(following page)*: Haplotype 1 gapless block length coverages. Annotations shown are (top-bottom) coding sequence (CDS), untranslated region (UTR), non exonic constrained elements (NXE), non genic constrained elements (NGE), CpG islands (island) and the union of Tandem Repeats Finder and RepeatMasker output (repeat). Red lines at the top of the tracks indicate the track goes off the scale and is visually clipped.
Assemblies are sorted in order of percent coverage, caluclated as the number of columns aligning to the haplotype divided by the length of the haplotype.
The color of the curve is indicative of the proportion of bases in that area of the genome that are in gapless blocks of a certain size threshold. The colored thresholds key is printed at the bottom. Feature positions have been discretized into 2,400 bins ($\sim$ 47 Kb/bin) for display.

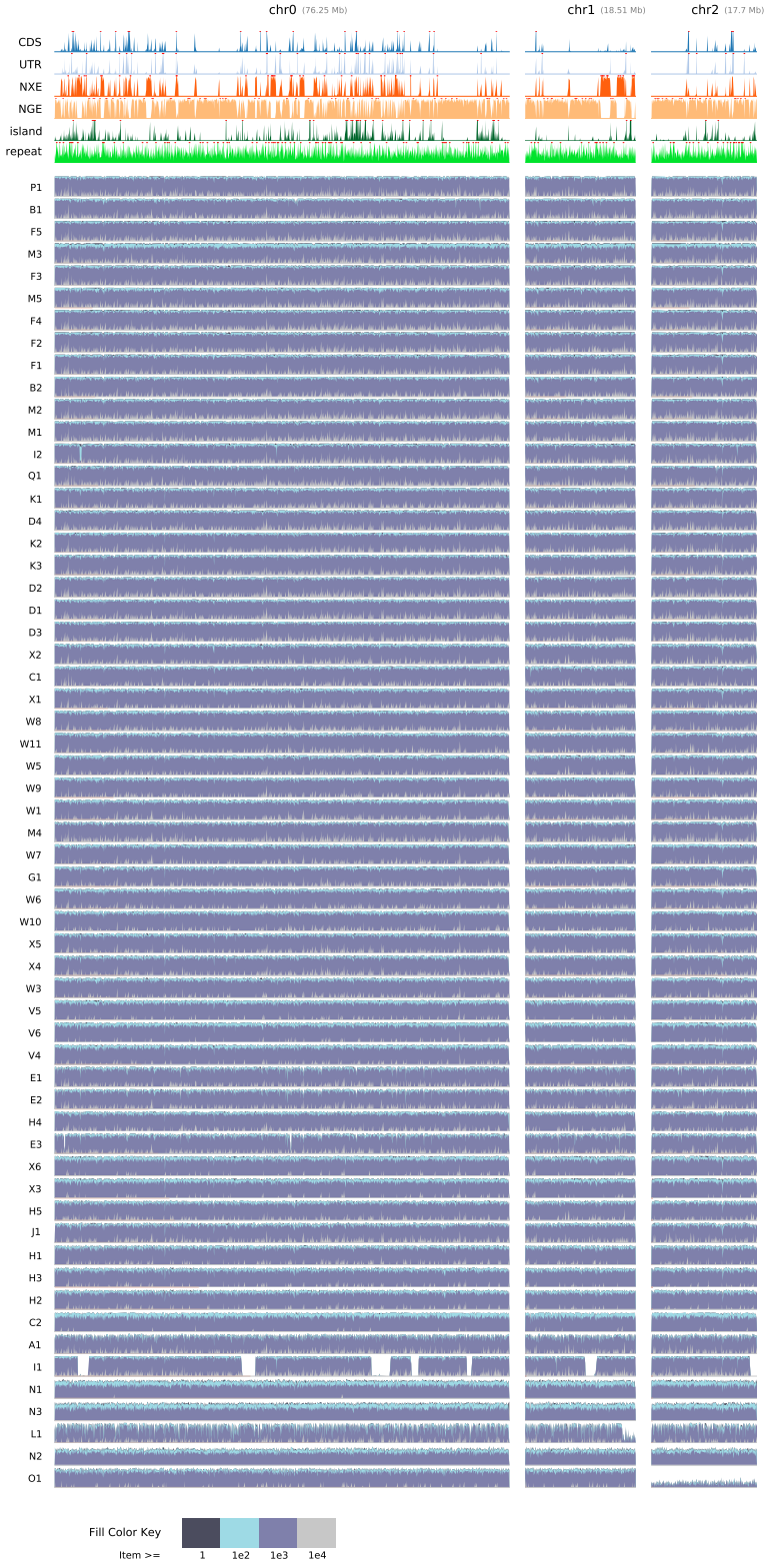

chr0 (76.25 Mb)
chr1 (18.51 Mb)
chr2 (17.7 Mb)
CDS
UTR
NXE
NGE
island
repeat
P1
B1
F5
M3
F3
M5
F4
F2
F1
B2
M2
M1
I2
Q1
K1
D4
K2
K3
D2
D1
D3
X2
C1
X1
W8
W11
W5
W9
W1
M4
W7
G1
W6
W10
X5
X4
W3
V5
V6
V4
E1
E2
H4
E3
X6
X3
H5
J1
H1
H3
H2
C2
A1
I1
N1
N3
L1
N2
O1
Fill Color Key
Item >=
1
1e2
1e3
1e4

Figure 3.12 *(following page)*: Haplotype 2 gapless block length coverages. Annotations shown are (top-bottom) coding sequence (CDS), untranslated region (UTR), non exonic constrained elements (NXE), non genic constrained elements (NGE), CpG islands (island) and the union of Tandem Repeats Finder and RepeatMasker output (repeat). Red lines at the top of the tracks indicate the track goes off the scale and is visually clipped.
Assemblies are sorted in order of percent coverage, caluclated as the number of columns aligning to the haplotype divided by the length of the haplotype.
The color of the curve is indicative of the proportion of bases in that area of the genome that are in gapless blocks of a certain size threshold. The colored thresholds key is printed at the bottom. Feature positions have been discretized into 2,400 bins ($\sim$ 47 Kb/bin) for display.


chr0 (76.29 Mb)
chr1 (18.54 Mb)
chr2 (17.71 Mb)
CDS
UTR
NXE
NGE
island
repeat
P1
B1
F5
M3
F3
M5
F4
F2
F1
B2
M2
M1
I2
Q1
K1
D4
K2
K3
D2
D1
D3
X2
C1
X1
W8
W11
W5
W9
W1
M4
W7
G1
W6
W10
X5
X4
W3
V5
V6
V4
E1
E2
H4
E3
X6
X3
H5
J1
H1
H3
H2
C2
A1
I1
N1
N3
L1
N2
O1
Fill Color Key
Item >=
1
1e2
1e3
1e4

Figure 3.13 *(following page)*: Aggregate stats plot, contigs.

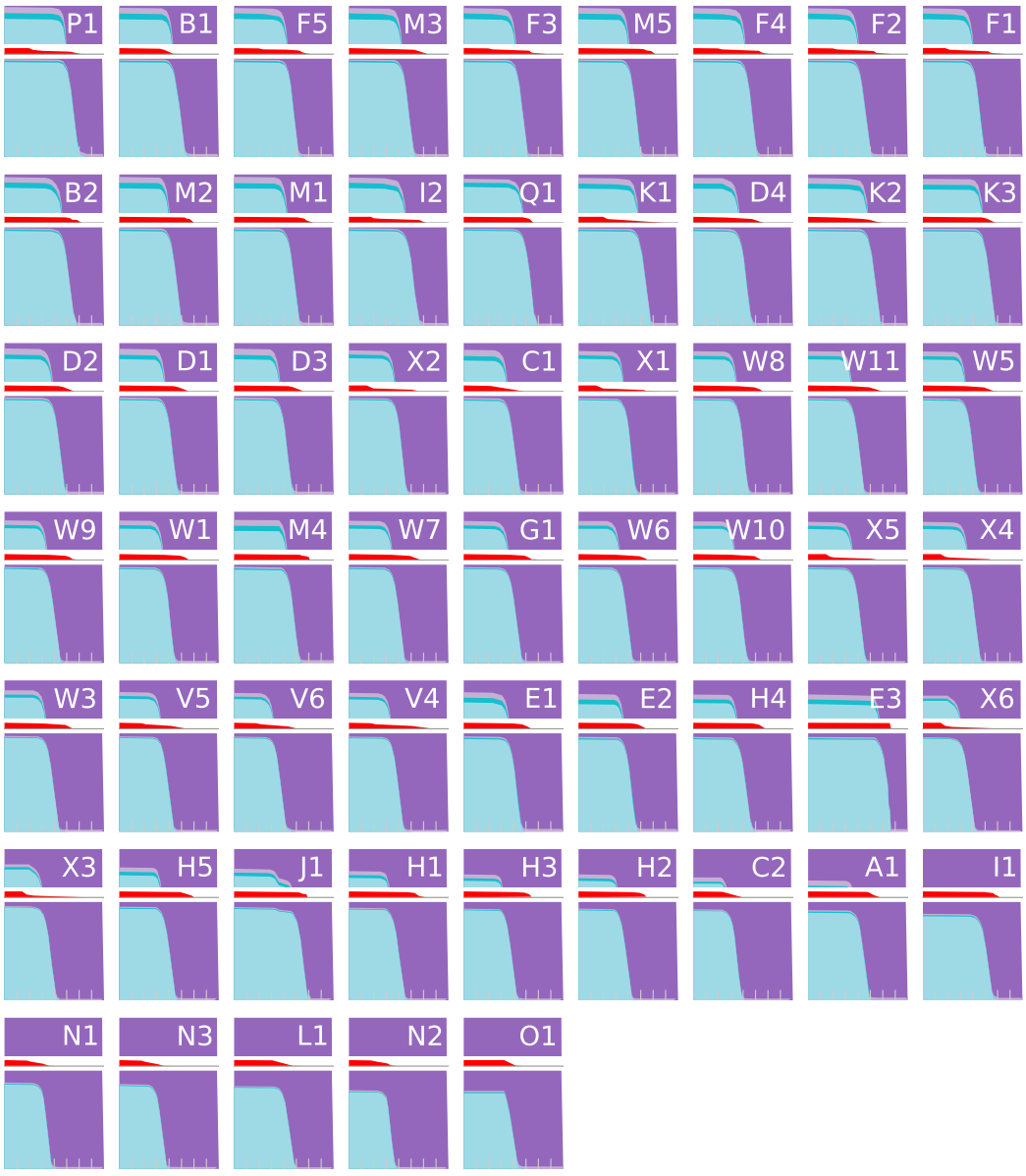


Figure 3.14 *(following page)*: Aggregate stats plot, scaffold paths.

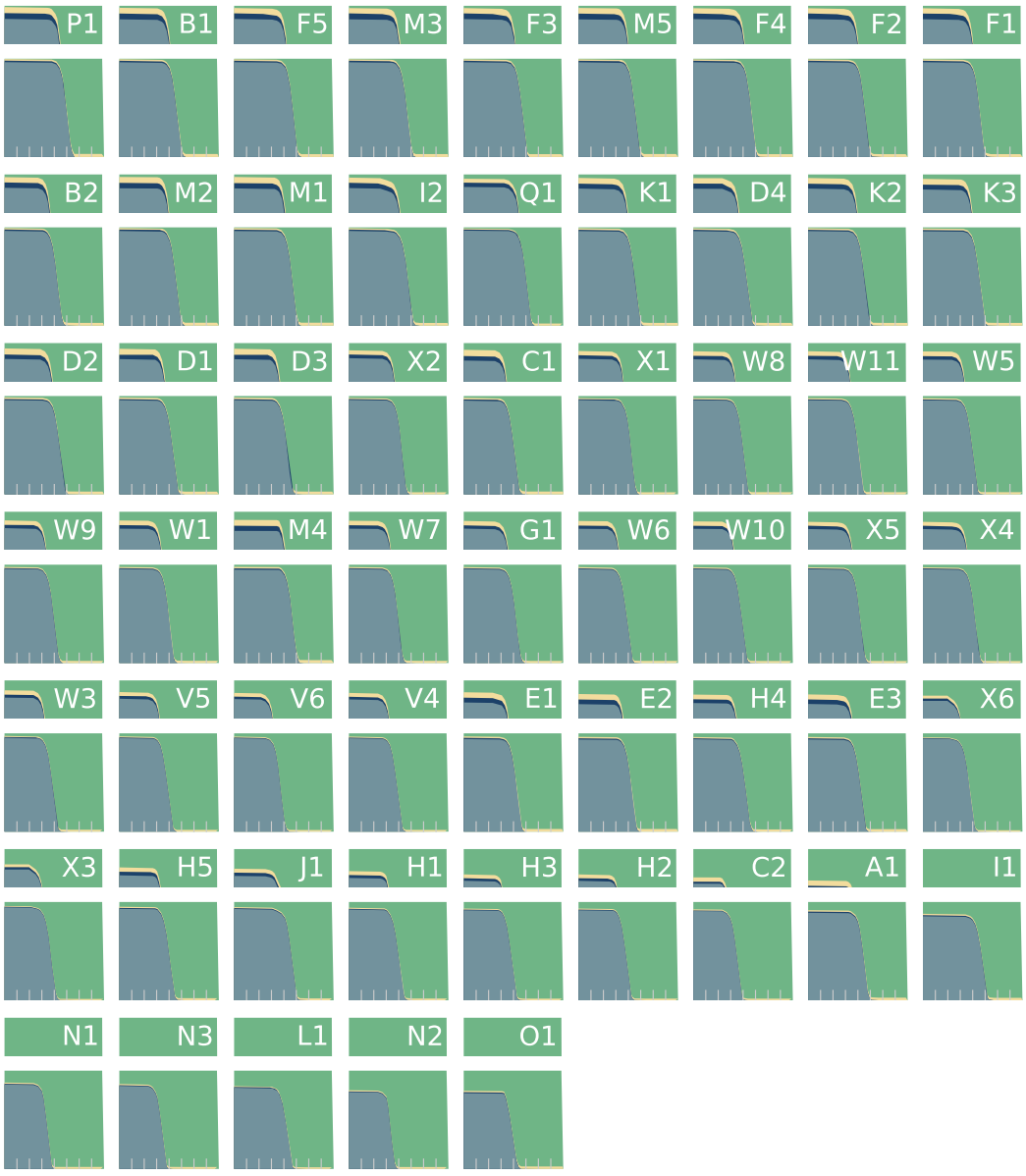


Figure 3.15 *(following page)*: Aggregate stats plot, haplotype paths.

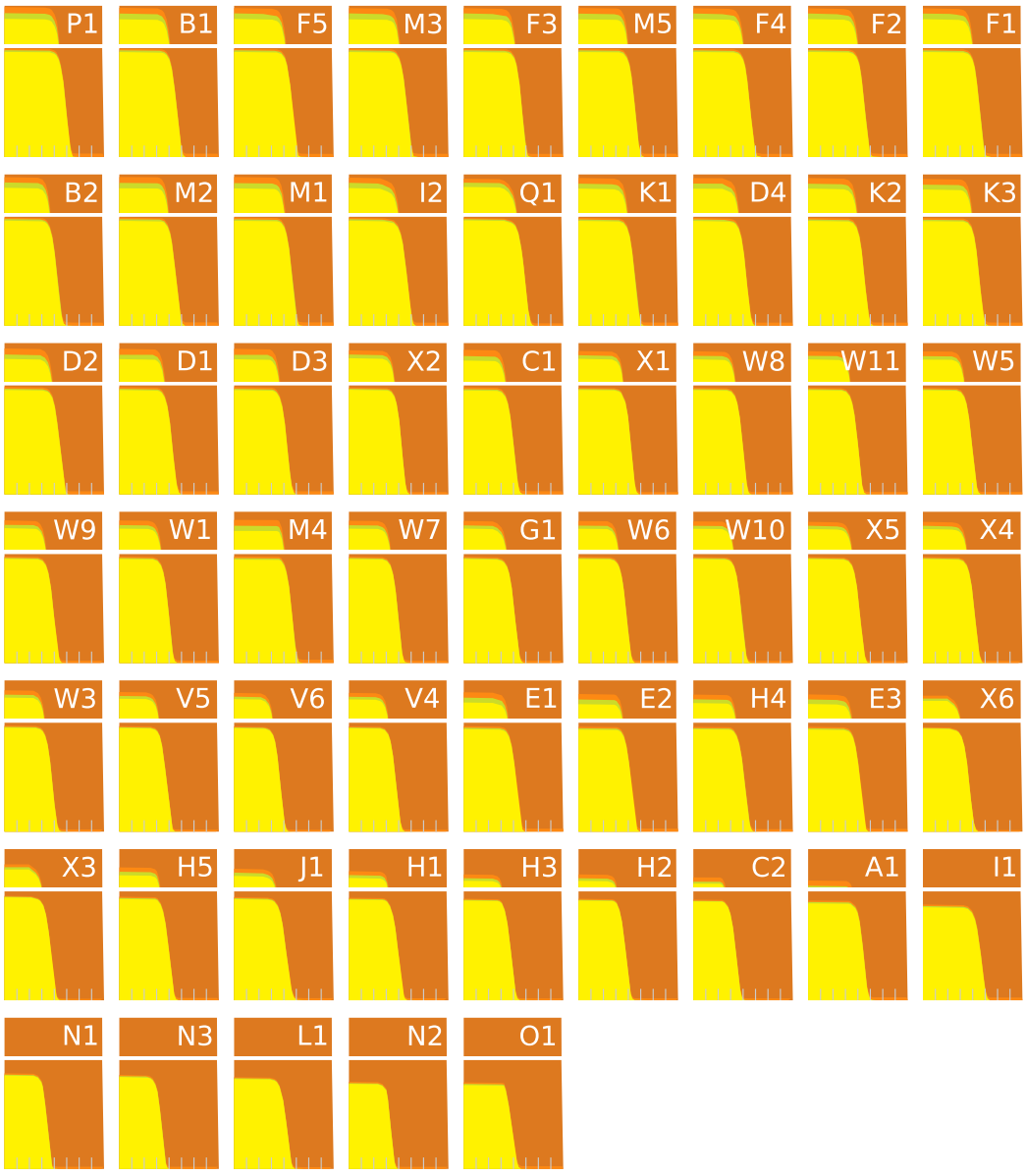


Figure 3.16 *(following page)*: Aggregate stats plot, blocks.

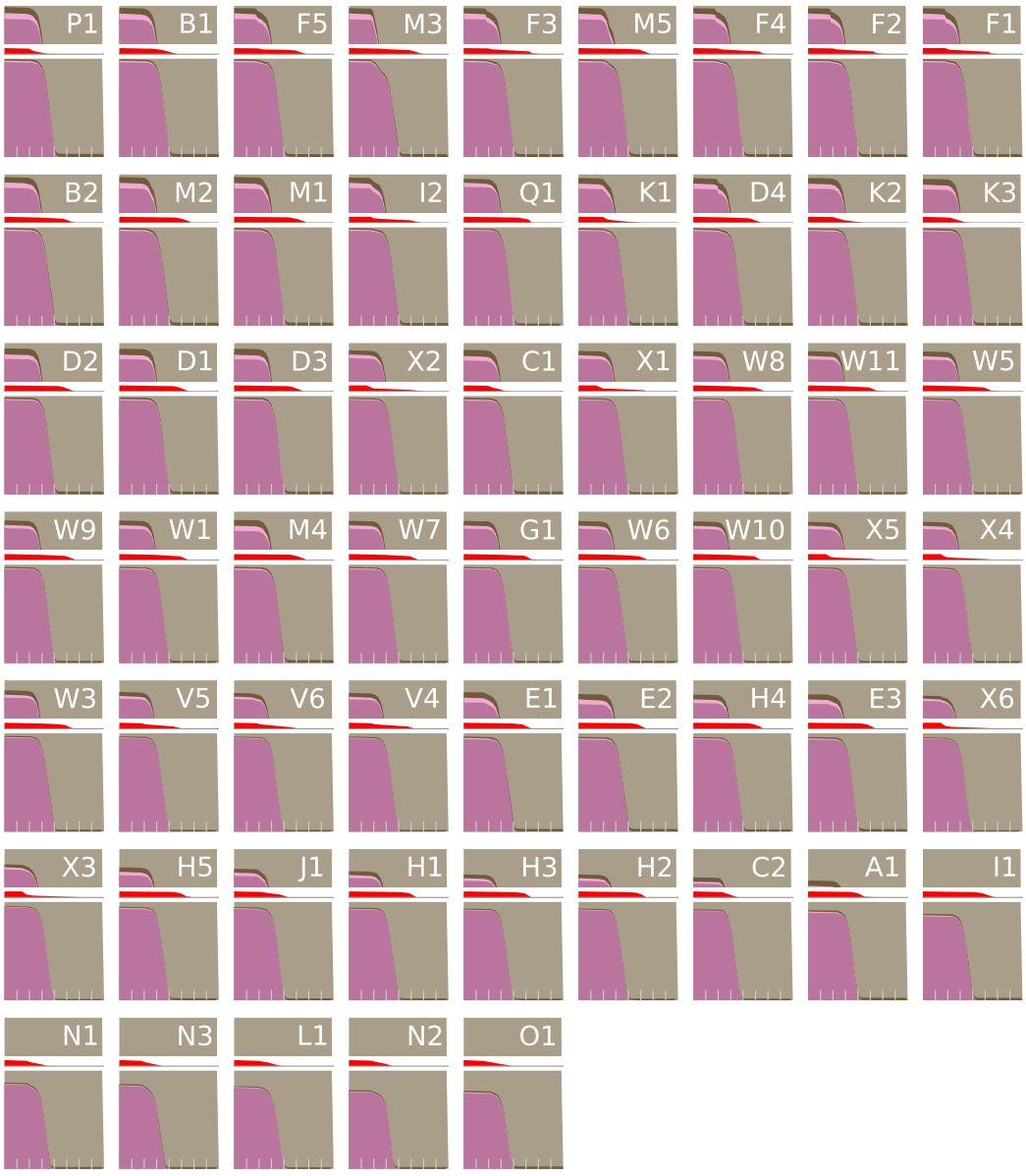


Figure 3.17 *(following page)*: SNP error plot, normalized.

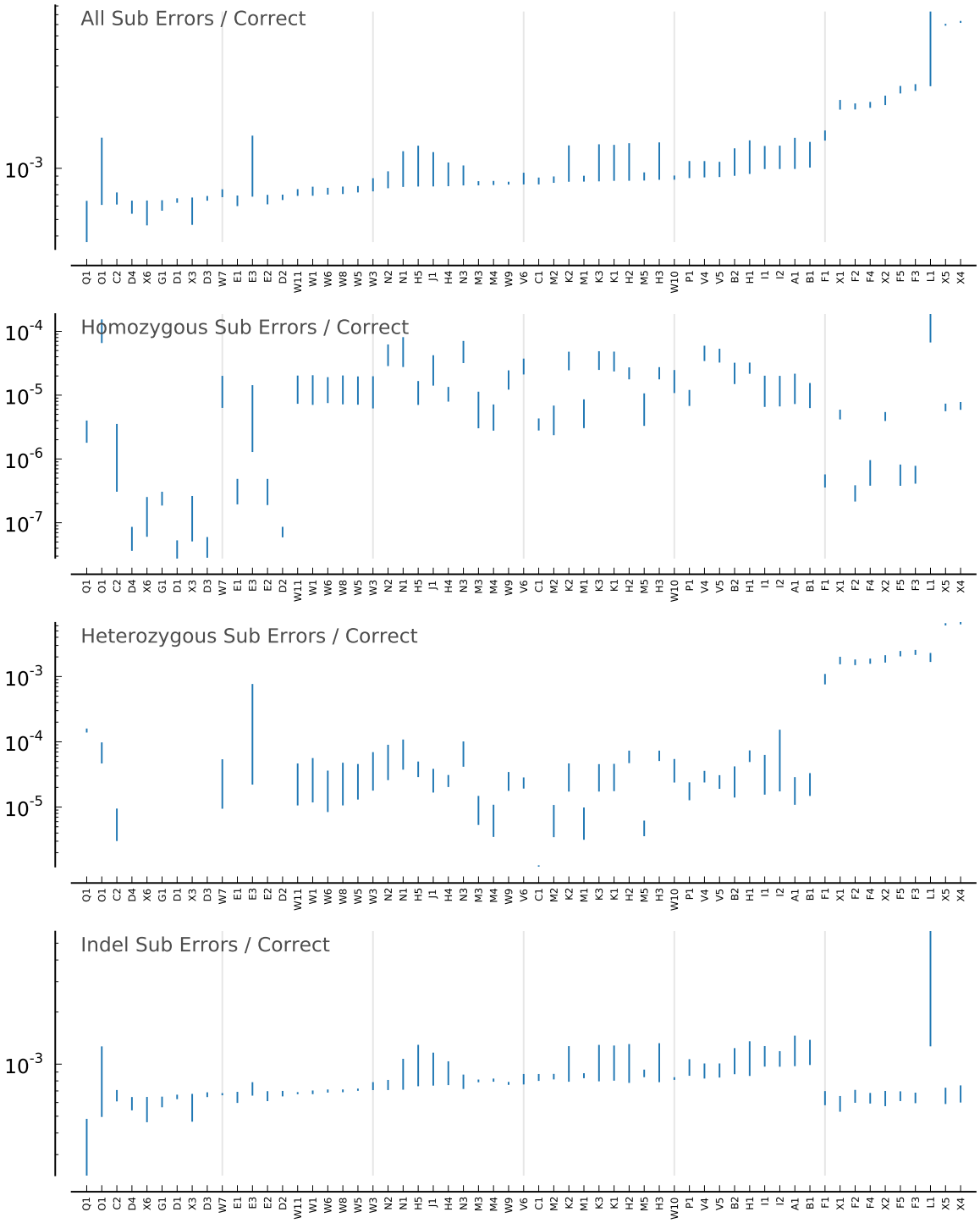


Figure 3.18 *(following page)*: N50 Statstics, as calculated in Table 3.2.

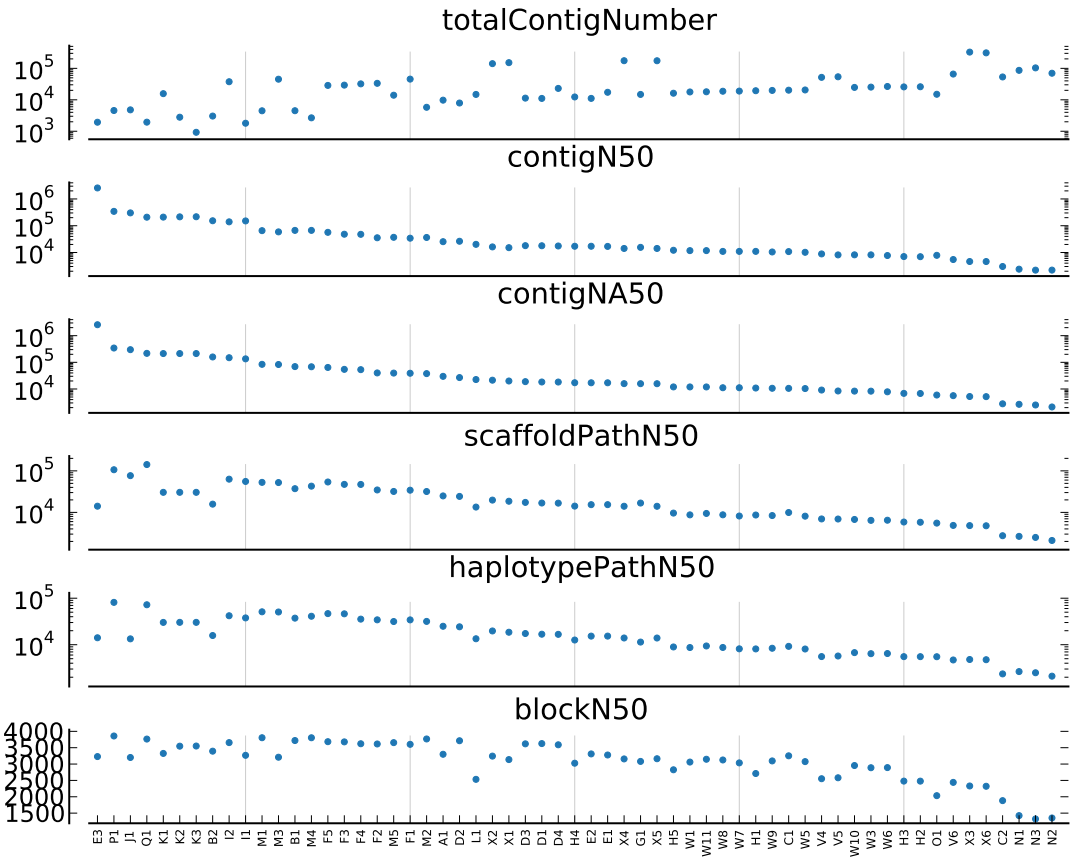


Table 3.2: N50 statistics.

ID	# Contigs	N50	NA50	SPA50	HPA50	BNA50
E3	1,937	**2,575,286**	**2,575,286**	14,105	14,105	3,229
P1	4,566	343,889	343,889	106,510	**81,390**	**3,858**
J1	4,791	301,691	298,640	76,698	13,367	3,201
Q1	1,946	208,256	217,104	**141,905**	72,336	3,764
K1	15,689	209,662	213,516	30,403	30,187	3,325
K2	2,796	214,562	213,516	30,431	30,273	3,547
K3	926	216,393	213,516	30,408	30,250	3,551
B2	3,040	153,374	158,964	15,796	15,766	3,394
I2	37,571	139,666	149,087	63,119	41,968	3,656
I1	1,798	151,121	135,002	55,591	37,703	3,267
M1	4,477	65,510	84,116	52,845	51,033	3,805
M3	45,200	58,916	82,867	52,321	50,529	3,209
B1	4,502	66,967	69,214	37,273	37,273	3,719
M4	2,672	67,017	68,146	42,862	40,946	3,804
F5	28,683	56,660	64,245	53,957	46,668	3,684
F3	29,300	48,200	54,510	47,299	46,169	3,678
F4	32,134	47,737	53,151	47,112	35,494	3,621
F2	33,437	35,510	40,038	34,694	34,277	3,614
M5	13,998	36,938	39,490	31,932	31,520	3,655
F1	45,487	34,247	38,946	34,249	34,197	3,603
M2	5,780	36,443	37,753	31,828	31,706	3,767
A1	9,741	25,383	29,791	24,930	24,930	3,299
D2	7,904	26,304	26,834	24,267	24,267	3,713
L1	14,822	20,165	22,620	13,434	13,434	2,533
X2	141,144	16,191	21,386	19,773	19,765	3,243
X1	151,852	15,165	19,938	18,562	18,548	3,137
D3	11,310	18,021	18,871	17,403	17,403	3,618
D1	11,067	17,882	18,300	16,819	16,819	3,626
D4	23,037	17,519	18,226	16,708	16,708	3,589
H4	12,316	17,117	17,186	14,172	12,616	3,026
E2	11,093	17,129	17,097	15,354	15,307	3,310
E1	17,341	16,897	17,087	15,367	15,316	3,277
X4	175,163	14,169	15,889	14,057	13,953	3,157
G1	14,817	15,585	15,814	16,787	11,396	3,081
X5	174,679	14,168	15,785	14,028	13,931	3,163
H5	16,190	12,130	11,963	9,620	8,946	2,825
W1	17,759	11,765	11,941	8,718	8,699	3,061
W11	17,979	11,777	11,930	9,432	9,396	3,147
W8	18,725	11,004	11,155	8,744	8,725	3,125
W7	18,929	11,013	11,136	8,178	8,157	3,036
H1	19,446	11,024	10,864	8,645	8,120	2,711
W9	19,862	10,472	10,639	8,394	8,394	3,096
C1	20,229	10,819	10,561	9,929	9,221	3,252
W5	20,561	10,181	10,388	8,109	8,087	3,075
V4	51,760	8,856	9,125	6,956	5,559	2,552
V5	53,949	8,141	8,419	6,924	5,715	2,581
W10	24,778	8,190	8,324	6,760	6,760	2,955
W3	25,328	8,169	8,292	6,425	6,415	2,888
W6	26,531	7,715	7,833	6,478	6,463	2,892
H3	25,709	7,057	6,849	5,861	5,540	2,479
H2	25,981	6,980	6,783	5,787	5,535	2,478
O1	14,994	7,847	5,928	5,518	5,518	2,034
V6	65,834	5,467	5,624	4,857	4,700	2,440
X3	**328,797**	4,611	5,217	4,808	4,808	2,330
X6	311,185	4,601	5,165	4,759	4,759	2,321
C2	53,273	3,003	2,782	2,744	2,357	1,882
N1	86,428	2,387	2,662	2,646	2,644	1,424
N3	103,555	2,196	2,512	2,496	2,494	1,322
N2	69,948	2,204	2,117	2,108	2,107	1,352

Table 3.3: Scaffold path statistics.

				Errors							
ID	hom/het sw	scf+bh scf	ctgE+ctgN	e-i	h-h	h-i	h-d	nonSpec	ePc	ePmb	$\sum$ err
D2	39,788	0	14,638	0.1154	8.00e-06	526	61	58	71	5	721
G1	41,122	4,582	25,862	0.0774	1.00e-05	702	63	41	89	58	953
A1	37,909	0	18,572	0.1562	1.50e-05	254	133	181	237	164	969
E2	48,462	50	20,443	0.1363	1.40e-05	735	80	194	92	43	1,144
D1	40,099	0	20,376	0.1247	1.30e-05	1,048	40	57	65	7	1,217
D3	41,472	0	20,844	0.1247	1.30e-05	1,068	36	57	70	15	1,246
E1	50,811	52	31,269	0.0943	1.50e-05	851	82	193	92	49	1,267
Q1	82,876	3,643	3,219	1.3746	2.40e-05	393	134	532	395	160	1,614
X1	74,913	28	31,052	0.0143	2.00e-05	1,177	25	269	142	112	1,725
D4	49,241	0	38,738	0.0860	1.80e-05	1,408	70	72	97	95	1,742
X2	76,072	29	28,346	0.0165	2.10e-05	1,263	37	252	192	100	1,844
X3	55,289	0	91,121	0.0058	1.80e-05	1,805	8	29	15	0	1,857
X6	52,096	0	91,517	0.0063	1.80e-05	1,807	8	31	34	2	1,882
F3	61,976	164	32,408	0.0847	2.20e-05	1,309	84	252	132	234	2,011
X4	56,175	115	31,623	0.0154	2.50e-05	1,298	43	318	307	62	2,028
X5	55,972	118	31,677	0.0155	2.50e-05	1,297	48	324	304	55	2,028
F1	61,605	18	40,563	0.0532	2.20e-05	1,481	75	196	115	163	2,030
P1	44,278	466	7,452	0.7792	3.20e-05	258	403	451	642	307	2,061
F2	60,853	145	38,906	0.0784	2.40e-05	1,484	84	241	133	220	2,162
M4	44,789	160	4,775	1.3177	3.20e-05	541	574	235	531	299	2,180
F5	62,272	598	31,181	0.0993	2.60e-05	1,237	127	296	177	407	2,244
M2	47,707	51	10,219	0.5637	2.90e-05	1,115	468	204	291	217	2,295
F4	61,431	975	36,378	0.1023	2.90e-05	1,522	128	320	196	475	2,641
C1	36,281	1,641	37,743	0.1881	3.50e-05	823	144	163	558	1,252	2,940
M1	61,331	208	7,922	1.2004	4.80e-05	820	955	395	622	608	3,400
I1	41,607	1,564	2,323	3.8615	7.10e-05	628	182	520	2,119	671	4,120
H4	42,504	1,143	20,952	0.4799	5.50e-05	2,489	114	553	992	100	4,248
B1	80,514	0	4,421	1.5657	6.30e-05	1,120	440	1,981	372	340	4,253
M5	59,302	281	25,201	0.4981	6.20e-05	1,425	1,218	381	899	549	4,472
C2	21,836	8,968	101,719	0.1158	6.00e-05	1,489	1,201	50	151	1,872	4,763
M3	100,287	251	83,011	0.1750	7.10e-05	2,017	1,080	582	910	748	5,337
N2	27,171	34	124,556	0.1033	8.10e-05	4,313	18	505	907	52	5,795
H1	39,608	1,118	32,605	0.4408	8.10e-05	3,030	118	978	1,570	208	5,904
I2	53,752	1,794	18,535	0.2725	9.20e-05	1,172	225	621	3,299	774	6,091
H5	41,266	1,104	27,364	0.5406	8.20e-05	3,760	116	986	1,345	96	6,303
W9	38,903	0	34,007	0.4946	9.00e-05	2,832	57	684	2,731	45	6,349
J1	45,040	11,240	5,494	1.7790	8.00e-05	925	493	498	1,119	3,375	6,410
K3	40,705	32	1,489	13.5194	1.13e-04	301	788	2,014	3,061	492	6,656
K2	41,437	32	3,517	4.5111	1.14e-04	399	793	2,012	3,056	490	6,750
H3	38,239	1,361	43,064	0.3636	8.90e-05	4,650	91	779	1,448	60	7,028
K1	42,894	32	20,604	0.8241	1.16e-04	680	789	2,001	3,067	534	7,071
H2	38,214	1,115	43,496	0.3623	9.00e-05	4,736	92	777	1,449	40	7,094
W10	39,261	0	42,641	0.4355	9.90e-05	3,387	32	643	3,014	26	7,102
V5	39,214	3,507	44,031	0.2273	1.13e-04	2,612	157	1,562	2,743	723	7,797
N3	35,462	149	176,550	0.1033	1.11e-04	6,388	66	728	1,258	204	8,644
W6	40,393	73	43,752	0.4906	1.19e-04	4,793	45	2,086	1,969	22	8,915
W8	41,600	77	29,657	0.7282	1.24e-04	4,379	53	2,376	2,179	40	9,027
W5	40,779	69	32,904	0.6643	1.25e-04	4,480	54	2,354	2,162	38	9,088
W11	41,602	86	28,146	0.7800	1.28e-04	4,534	59	2,409	2,256	39	9,297
V6	39,838	907	62,312	0.1764	1.07e-04	4,871	303	791	1,185	2,182	9,332
V4	37,855	4,500	41,115	0.3009	1.44e-04	2,383	764	1,635	2,811	2,768	10,361
W1	41,752	90	27,615	0.9291	1.51e-04	4,270	62	3,156	2,880	32	10,400
W7	41,500	75	29,655	0.8661	1.50e-04	4,450	45	3,079	2,834	27	10,435
W3	40,473	51	40,978	0.6628	1.54e-04	5,302	40	2,878	2,809	28	11,057
O1	35,588	0	23,857	1.0870	1.84e-04	6,101	1,365	1,260	2,284	377	11,387
N1	38,388	142	147,175	0.1659	1.46e-04	8,428	71	995	1,806	164	11,464
E3	49,477	0	3,544	13.4311	2.41e-04	67	511	2,296	9,963	408	13,245
B2	50,879	32	4,874	7.6536	2.09e-04	637	783	5,398	1,959	6,348	15,125
L1	49,260	0	22,579	2.8928	4.55e-04	5,789	2,103	6,781	6,859	5,602	27,134

Table 3.4: Subsitution statistics table, Homozygous class.

Assembly	Total	Calls	Correct (bits)	Errors
P1	107,821,484 – 108,416,677	107,475,470 – 108,037,854	214,947,712.00 – 216,069,946.00	1,500 – 2,532
B1	108,960,814 – 109,581,319	108,388,158 – 108,973,477	216,773,514.00 – 217,940,394.00	1,401 – 3,280
F5	97,818,962 – 98,365,958	97,192,029 – 97,699,187	194,382,777.00 – 195,395,532.00	76 – 155
M3	71,551,605 – 71,969,021	70,904,442 – 71,275,640	141,807,836.00 – 142,547,308.00	442 – 1,568
F3	97,803,309 – 98,352,965	97,156,865 – 97,664,777	194,312,828.00 – 195,328,173.00	82 – 148
M5	104,581,692 – 105,149,319	103,977,385 – 104,499,826	207,952,908.00 – 208,993,426.00	707 – 2,164
F4	100,042,257 – 100,613,611	99,368,372 – 99,897,240	198,735,383.00 – 199,790,857.00	78 – 186
F2	100,034,692 – 100,609,131	99,348,767 – 99,878,337	198,696,942.00 – 199,755,641.00	44 – 75
F1	99,990,258 – 100,565,939	99,304,668 – 99,834,763	198,608,578.42 – 199,668,465.42	73 – 111
B2	108,028,121 – 108,761,484	107,315,720 – 107,958,959	214,624,640.00 – 215,903,846.00	3,299 – 6,754
M2	108,034,382 – 108,626,598	107,329,830 – 107,870,867	214,658,544.00 – 215,738,488.00	521 – 1,438
M1	79,070,669 – 79,512,246	78,367,042 – 78,759,434	156,732,786.00 – 157,515,016.00	491 – 1,314
I2	108,081,096 – 108,814,321	107,221,079 – 107,873,260	214,436,520.00 – 215,728,370.00	1,471 – 4,204
Q1	105,634,099 – 106,191,597	104,400,809 – 104,915,082	208,796,220.00 – 209,820,946.00	386 – 815
K1	107,089,490 – 107,970,124	106,231,521 – 106,945,316	212,452,692.00 – 213,870,430.00	5,150 – 9,996
D4	107,999,138 – 108,677,636	107,013,548 – 107,600,740	214,027,080.00 – 215,201,444.00	8 – 18
K2	108,550,808 – 109,349,358	107,670,972 – 108,302,410	215,331,006.00 – 216,584,434.00	5,444 – 10,088
K3	108,611,132 – 109,385,076	107,458,620 – 108,075,269	214,906,182.00 – 216,129,838.00	5,503 – 10,245
D2	108,141,752 – 108,865,817	106,846,982 – 107,352,006	213,693,938.00 – 214,703,976.00	13 – 18
D1	107,772,915 – 108,544,819	106,406,177 – 106,922,341	212,812,342.00 – 213,844,660.00	6 – 11
D3	104,350,290 – 105,105,307	102,997,160 – 103,498,102	205,994,308.00 – 206,996,180.00	6 – 12
X2	79,239,233 – 79,990,966	77,382,200 – 77,871,325	154,762,522.83 – 155,739,569.25	628 – 822
C1	107,995,683 – 109,484,691	106,928,115 – 107,548,512	213,851,312.00 – 215,077,024.00	615 – 900
X1	80,406,978 – 81,196,668	78,360,674 – 78,873,243	156,719,343.66 – 157,743,082.66	672 – 907
W8	109,193,708 – 110,234,437	107,139,516 – 107,848,948	214,252,932.00 – 215,653,668.00	1,583 – 4,262
W11	109,194,772 – 110,214,561	107,129,828 – 107,831,708	214,235,328.00 – 215,619,862.00	1,616 – 4,246
W5	109,123,615 – 110,209,848	107,044,765 – 107,765,270	214,062,756.00 – 215,485,536.00	1,562 – 4,113
W9	109,213,229 – 110,232,757	107,044,822 – 107,701,039	214,084,260.00 – 215,391,840.00	2,692 – 5,119
W1	109,163,268 – 110,224,825	106,909,158 – 107,648,097	213,786,064.00 – 215,244,720.00	1,551 – 4,294
M4	108,784,300 – 109,379,247	106,407,542 – 106,944,439	212,813,426.00 – 213,884,110.00	609 – 1,484
W7	109,177,431 – 110,257,267	106,877,032 – 107,609,778	213,724,318.00 – 215,170,768.00	1,390 – 4,193
G1	109,156,013 – 110,346,770	106,984,207 – 107,681,757	213,955,608.00 – 215,310,228.00	41 – 64
W6	109,080,824 – 110,244,960	106,810,762 – 107,547,714	213,601,756.00 – 215,057,646.00	1,654 – 4,000
W10	109,152,890 – 110,261,108	106,751,246 – 107,445,597	213,497,746.00 – 214,880,856.00	2,373 – 5,169
X5	105,280,830 – 106,319,236	102,682,740 – 103,354,269	205,358,873.91 – 206,699,207.74	1,185 – 1,478
X4	104,969,117 – 106,005,587	102,366,017 – 103,037,469	204,725,177.49 – 206,065,327.91	1,243 – 1,561
W3	109,146,021 – 110,312,968	106,667,991 – 107,413,680	213,319,492.00 – 214,791,202.00	1,361 – 4,127
V5	108,872,432 – 110,364,465	106,062,526 – 106,930,269	212,103,782.00 – 213,791,104.00	7,085 – 11,086
V6	108,920,310 – 110,354,787	105,769,637 – 106,667,798	211,528,070.00 – 213,302,438.00	4,598 – 7,745
V4	108,757,572 – 110,363,899	105,765,024 – 106,658,751	211,502,070.00 – 213,213,864.00	7,466 – 12,342
E1	108,234,018 – 109,183,195	105,036,662 – 105,666,068	210,071,100.00 – 211,328,142.00	42 – 100
E2	108,756,923 – 109,785,776	105,167,084 – 105,765,087	210,331,968.00 – 211,526,330.00	41 – 100
H4	108,068,300 – 109,108,408	104,214,426 – 104,897,436	208,421,418.00 – 209,773,994.00	1,700 – 2,742
E3	108,775,955 – 109,854,286	104,968,462 – 105,708,062	209,936,368.00 – 211,410,236.00	278 – 2,944
X6	108,437,056 – 110,212,352	104,681,739 – 105,744,241	209,363,452.00 – 211,488,378.00	13 – 52
X3	108,463,072 – 110,237,427	104,701,577 – 105,757,697	209,403,132.00 – 211,515,286.00	11 – 54
H5	108,738,853 – 109,900,362	103,691,222 – 104,459,524	207,376,006.00 – 208,897,476.00	1,502 – 3,378
J1	109,307,624 – 110,037,419	103,578,656 – 104,192,856	207,132,892.00 – 208,305,362.00	3,010 – 8,540
H1	108,583,668 – 109,825,235	102,511,065 – 103,290,776	205,011,450.00 – 206,559,850.00	4,571 – 6,496
H3	108,924,626 – 110,389,766	102,026,137 – 102,891,606	204,040,992.00 – 205,756,184.00	3,718 – 5,486
H2	108,909,616 – 110,376,813	102,013,172 – 102,880,000	204,015,082.00 – 205,732,812.00	3,718 – 5,444
C2	107,032,299 – 110,591,207	101,422,736 – 102,324,638	202,817,188.00 – 204,485,796.00	64 – 702
A1	89,278,573 – 89,999,824	80,139,373 – 80,644,836	160,276,348.00 – 161,282,878.00	1,199 – 3,397
I1	109,274,593 – 109,998,671	95,859,650 – 96,440,001	191,713,066.00 – 192,858,436.00	1,292 – 3,779
N1	80,315,755 – 83,700,670	68,087,614 – 70,113,684	136,124,878.00 – 140,127,048.00	3,872 – 11,050
N3	76,776,616 – 80,970,919	63,246,902 – 65,763,077	126,410,052.00 – 131,381,156.00	4,143 – 9,032
L1	98,281,517 – 99,262,240	80,820,755 – 81,581,285	161,619,208.00 – 163,101,680.00	11,151 – 30,441
N2	92,565,231 – 95,104,910	71,389,031 – 72,748,731	142,751,558.00 – 145,453,938.00	4,199 – 8,802
O1	108,045,903 – 108,951,478	85,158,225 – 85,811,948	170,293,418.00 – 171,572,566.00	11,516 – 25,665

Table 3.5: Subsitution statistics table, Heterozygous class.

Assembly	Total	Calls	Correct (bits)	Errors
P1	421,708 – 432,050	419,968 – 429,848	839,904.00 – 859,632.00	11 – 20
B1	428,748 – 437,625	426,068 – 434,666	852,110.00 – 869,276.00	13 – 28
F5	381,627 – 389,289	378,382 – 385,345	753,520.00 – 766,820.00	1,580 – 1,837
M3	277,102 – 281,107	274,023 – 277,682	548,038.00 – 555,340.00	3 – 8
F3	381,474 – 389,288	378,136 – 385,060	752,884.00 – 766,166.00	1,662 – 1,906
M5	410,098 – 417,440	407,276 – 414,269	814,540.00 – 828,518.00	3 – 5
F4	389,681 – 397,504	386,235 – 393,291	769,900.00 – 783,534.00	1,245 – 1,433
F2	389,653 – 397,558	386,121 – 392,964	769,830.00 – 783,076.00	1,186 – 1,390
F1	389,205 – 397,277	385,682 – 392,637	770,084.42 – 783,488.42	601 – 834
B2	419,996 – 433,136	416,349 – 428,442	832,670.00 – 856,792.00	12 – 35
M2	424,871 – 433,450	421,552 – 429,718	843,098.00 – 859,416.00	3 – 9
M1	311,166 – 317,542	307,843 – 313,823	615,678.00 – 627,622.00	2 – 6
I2	423,181 – 432,199	419,081 – 426,660	838,118.00 – 853,016.00	15 – 127
Q1	415,374 – 425,609	409,789 – 418,816	160,952.00 – 173,924.00	23 – 27
K1	419,641 – 431,126	415,464 – 425,503	830,896.00 – 850,928.00	15 – 38
D4	421,001 – 432,419	416,034 – 423,813	832,068.00 – 847,626.00	0 – 0
K2	426,211 – 436,207	421,963 – 430,494	843,894.00 – 860,908.00	15 – 39
K3	426,783 – 436,632	421,408 – 429,919	842,784.00 – 859,760.00	15 – 38
D2	421,157 – 434,767	413,511 – 419,945	827,022.00 – 839,890.00	0 – 0
D1	419,027 – 433,432	411,097 – 417,224	822,194.00 – 834,448.00	0 – 0
D3	405,710 – 419,981	397,846 – 403,850	795,692.00 – 807,700.00	0 – 0
X2	305,272 – 319,306	295,941 – 302,179	589,799.89 – 601,727.89	994 – 1,240
C1	410,716 – 432,828	401,911 – 408,634	803,492.00 – 816,264.00	1 – 1
X1	309,868 – 324,469	299,589 – 305,889	597,192.23 – 609,240.23	947 – 1,190
W8	422,545 – 439,958	412,279 – 419,949	824,492.00 – 839,722.00	9 – 39
W11	422,826 – 439,880	412,509 – 420,118	824,954.00 – 840,056.00	9 – 38
W5	420,117 – 439,815	409,688 – 416,978	819,292.00 – 833,762.00	11 – 37
W9	422,097 – 439,967	411,322 – 417,889	822,614.00 – 835,722.00	15 – 28
W1	423,544 – 439,925	412,374 – 420,218	824,648.00 – 840,216.00	10 – 46
M4	428,115 – 436,978	417,966 – 426,307	835,920.00 – 852,580.00	3 – 9
W7	423,368 – 440,084	411,875 – 419,496	823,648.00 – 838,770.00	8 – 44
G1	413,051 – 441,091	401,549 – 411,073	800,976.00 – 812,732.00	0 – 0
W6	416,281 – 439,697	404,995 – 412,381	809,934.00 – 824,596.00	7 – 29
W10	420,179 – 439,680	408,350 – 415,325	816,660.00 – 830,562.00	20 – 44
X5	409,874 – 423,099	396,971 – 403,405	783,980.83 – 796,343.25	4,894 – 5,087
X4	408,857 – 421,961	395,949 – 402,310	781,571.00 – 793,751.83	5,071 – 5,283
W3	419,853 – 439,919	407,565 – 415,051	815,054.00 – 829,906.00	15 – 56
V5	422,036 – 440,269	408,773 – 418,162	817,274.00 – 835,204.00	16 – 25
V6	419,615 – 439,603	404,781 – 415,701	809,462.00 – 830,846.00	16 – 23
V4	422,165 – 440,464	407,856 – 417,220	814,982.00 – 831,926.00	20 – 29
E1	408,595 – 429,530	394,263 – 403,066	785,554.00 – 802,530.00	0 – 2
E2	413,233 – 438,639	395,795 – 404,142	788,636.00 – 804,734.00	0 – 2
H4	424,958 – 435,871	407,517 – 415,375	814,802.00 – 830,070.00	17 – 25
E3	412,621 – 439,130	395,697 – 409,394	791,358.00 – 817,566.00	18 – 611
X6	412,330 – 436,617	394,658 – 401,317	789,316.00 – 802,634.00	0 – 0
X3	412,504 – 436,689	394,809 – 401,023	789,618.00 – 802,046.00	0 – 0
H5	427,347 – 439,318	404,349 – 412,998	808,458.00 – 825,312.00	24 – 40
J1	432,048 – 439,528	407,486 – 414,252	814,786.00 – 828,080.00	14 – 31
H1	424,863 – 438,424	396,066 – 404,568	791,946.00 – 808,564.00	40 – 58
H3	424,776 – 440,926	391,811 – 400,821	783,282.00 – 800,804.00	41 – 57
H2	424,731 – 440,953	391,759 – 400,831	783,186.00 – 800,830.00	38 – 57
C2	377,060 – 441,719	323,830 – 332,130	643,768.00 – 650,916.00	2 – 6
A1	356,848 – 367,886	313,738 – 321,827	627,462.00 – 643,618.00	7 – 18
I1	429,957 – 439,141	376,063 – 383,892	752,078.00 – 767,588.00	12 – 47
N1	305,379 – 339,277	247,666 – 262,788	493,466.00 – 521,994.00	19 – 55
N3	297,319 – 331,460	234,954 – 250,468	468,150.00 – 497,438.00	20 – 49
L1	387,889 – 396,512	317,518 – 324,673	632,856.00 – 646,458.00	1,090 – 1,444
N2	356,453 – 386,257	262,951 – 274,546	525,602.00 – 548,516.00	14 – 48
O1	425,297 – 434,585	333,807 – 341,427	667,550.00 – 682,724.00	32 – 65

Table 3.6: Subsitution statistics table, Indel class.

Assembly	Total	Calls	Correct (bits)	Errors
P1	3,155,268 – 3,540,700	1,521,634 – 1,716,959	3,038,006.00 – 3,426,412.00	2,630 – 3,620
B1	3,053,841 – 3,466,130	1,440,451 – 1,749,429	2,875,154.00 – 3,489,322.00	2,874 – 4,768
F5	2,767,272 – 3,164,605	1,261,291 – 1,566,319	2,518,988.00 – 3,127,222.00	1,735 – 1,934
M3	2,624,712 – 2,995,427	1,074,161 – 1,408,975	2,144,914.00 – 2,813,152.00	1,704 – 2,266
F3	2,779,611 – 3,178,046	1,276,989 – 1,583,179	2,550,377.00 – 3,161,593.00	1,726 – 1,901
M5	3,002,862 – 3,407,750	1,433,840 – 1,708,030	2,862,804.00 – 3,409,238.00	2,433 – 3,135
F4	2,772,485 – 3,173,973	1,274,187 – 1,583,172	2,544,716.00 – 3,160,763.00	1,716 – 1,892
F2	2,758,123 – 3,162,489	1,272,405 – 1,584,568	2,541,202.00 – 3,164,760.00	1,784 – 1,913
F1	2,793,458 – 3,193,355	1,280,590 – 1,588,559	2,557,560.00 – 3,172,521.00	1,767 – 1,857
B2	2,732,421 – 3,188,445	1,280,179 – 1,594,481	2,555,814.00 – 3,181,080.00	2,259 – 3,894
M2	2,997,904 – 3,390,024	1,401,048 – 1,614,139	2,797,478.00 – 3,222,540.00	2,307 – 2,800
M1	2,746,787 – 3,080,879	1,070,026 – 1,224,885	2,136,464.00 – 2,445,266.00	1,791 – 2,143
I2	2,701,965 – 3,143,423	1,249,564 – 1,577,816	2,494,016.00 – 3,144,390.00	2,436 – 3,696
Q1	2,060,212 – 2,423,335	937,734 – 1,245,544	1,871,056.00 – 2,483,110.00	426 – 1,186
K1	2,772,237 – 3,164,218	1,231,746 – 1,434,638	2,459,500.00 – 2,862,004.00	1,995 – 3,626
D4	2,791,469 – 3,187,403	1,275,033 – 1,548,458	2,546,814.00 – 3,093,534.00	1,626 – 1,691
K2	2,793,304 – 3,185,337	1,266,635 – 1,459,716	2,529,222.00 – 2,912,094.00	2,023 – 3,659
K3	2,794,782 – 3,185,607	1,263,667 – 1,440,996	2,523,274.00 – 2,874,628.00	2,029 – 3,672
D2	2,772,129 – 3,143,607	1,213,099 – 1,314,510	2,422,842.00 – 2,625,564.00	1,678 – 1,728
D1	2,705,290 – 3,077,308	1,167,503 – 1,269,482	2,331,934.00 – 2,535,748.00	1,536 – 1,608
D3	2,733,071 – 3,103,547	1,168,770 – 1,269,450	2,334,366.00 – 2,535,592.00	1,587 – 1,654
X2	1,840,656 – 2,166,723	700,225 – 909,364	1,398,424.00 – 1,815,350.42	968 – 1,049
C1	2,826,068 – 3,222,508	1,212,471 – 1,347,183	2,420,966.00 – 2,685,340.00	1,958 – 2,329
X1	1,801,359 – 2,129,752	654,492 – 864,407	1,307,197.00 – 1,725,831.00	846 – 926
W8	2,044,907 – 2,420,601	882,640 – 1,029,262	1,762,628.00 – 2,055,006.00	1,226 – 1,451
W11	2,069,244 – 2,446,102	899,019 – 1,046,181	1,795,430.00 – 2,088,912.00	1,217 – 1,423
W5	2,070,348 – 2,447,854	895,766 – 1,028,800	1,788,828.00 – 2,054,078.00	1,268 – 1,467
W9	1,955,533 – 2,328,994	841,107 – 969,162	1,679,592.00 – 1,935,354.00	1,311 – 1,485
W1	2,075,651 – 2,453,341	905,380 – 1,058,778	1,808,046.00 – 2,113,630.00	1,226 – 1,472
M4	2,910,014 – 3,294,504	1,349,054 – 1,542,590	2,693,794.00 – 3,080,006.00	2,155 – 2,514
W7	2,005,360 – 2,382,446	877,397 – 1,029,185	1,752,130.00 – 2,054,466.00	1,170 – 1,382
G1	1,933,605 – 2,304,180	813,626 – 970,831	1,624,880.00 – 1,928,602.00	1,043 – 1,098
W6	1,958,719 – 2,334,254	822,114 – 946,549	1,641,882.00 – 1,890,034.00	1,137 – 1,334
W10	1,852,801 – 2,224,890	774,357 – 898,074	1,546,144.00 – 1,793,208.00	1,285 – 1,470
X5	1,768,904 – 2,134,099	729,569 – 964,024	1,456,897.42 – 1,924,004.83	1,053 – 1,144
X4	1,766,601 – 2,132,972	718,895 – 954,346	1,435,522.42 – 1,904,576.83	1,070 – 1,155
W3	1,809,799 – 2,184,873	762,257 – 887,975	1,522,216.00 – 1,772,740.00	1,089 – 1,380
V5	1,607,768 – 1,999,314	657,490 – 833,806	1,312,662.00 – 1,658,304.00	1,110 – 1,655
V6	1,451,820 – 1,841,963	560,946 – 735,902	1,120,122.00 – 1,467,246.00	864 – 1,273
V4	1,551,498 – 1,938,567	621,878 – 789,930	1,241,562.00 – 1,568,298.00	1,036 – 1,565
E1	2,676,655 – 3,062,217	1,205,446 – 1,443,335	2,407,452.00 – 2,881,972.00	1,647 – 1,739
E2	2,662,361 – 3,044,670	1,191,904 – 1,404,952	2,380,370.00 – 2,805,176.00	1,643 – 1,733
H4	2,161,999 – 2,535,782	949,061 – 1,123,141	1,895,060.00 – 2,240,472.00	1,449 – 2,301
E3	2,488,705 – 2,873,311	1,146,992 – 1,363,595	2,290,938.00 – 2,722,954.00	1,523 – 2,118
X6	1,185,234 – 1,558,748	455,655 – 691,637	910,146.00 – 1,381,984.00	582 – 645
X3	1,158,499 – 1,529,620	443,668 – 688,821	886,156.00 – 1,376,348.00	590 – 647
H5	1,997,393 – 2,372,708	887,453 – 1,060,393	1,772,218.00 – 2,114,642.00	1,335 – 2,704
J1	1,771,527 – 2,157,718	757,261 – 983,039	1,511,650.00 – 1,956,348.00	1,148 – 2,255
H1	1,659,874 – 2,033,050	701,976 – 862,178	1,401,520.00 – 1,719,398.00	1,211 – 2,305
H3	1,530,034 – 1,901,797	617,914 – 770,538	1,233,808.00 – 1,536,504.00	981 – 2,008
H2	1,549,284 – 1,921,180	626,343 – 778,688	1,250,642.00 – 1,552,804.00	985 – 2,006
C2	1,506,313 – 1,874,511	549,592 – 593,180	1,097,682.00 – 1,165,892.00	676 – 817
A1	2,553,636 – 2,918,017	1,178,640 – 1,316,853	2,352,650.00 – 2,626,114.00	2,315 – 3,796
I1	2,418,790 – 2,819,795	1,072,681 – 1,268,760	2,140,966.00 – 2,527,462.00	2,094 – 3,178
N1	1,752,542 – 2,129,827	566,357 – 703,522	1,130,822.00 – 1,400,404.00	813 – 1,487
N3	1,679,463 – 2,052,370	530,923 – 653,535	1,059,928.00 – 1,301,104.00	770 – 1,120
L1	1,390,045 – 1,789,433	558,502 – 761,013	1,114,150.00 – 1,504,468.00	1,427 – 8,778
N2	1,708,456 – 2,080,270	590,343 – 693,819	1,178,962.00 – 1,384,798.00	844 – 1,110
O1	2,405,864 – 2,797,300	1,038,036 – 1,202,358	2,073,996.00 – 2,398,708.00	1,038 – 3,004

3.2 Individual results

3.2.1 A, GIS_CMB1

Affiliation: Agency for Science, Technology and Research, Singapore
Contact: Pramila Ariyaratne
Software: **PE-Assembler**
Number of entries: 1

ID	Total	Hap 1	Hap 2	Bac
A1	0.90867	0.90879	0.90854	0.99971

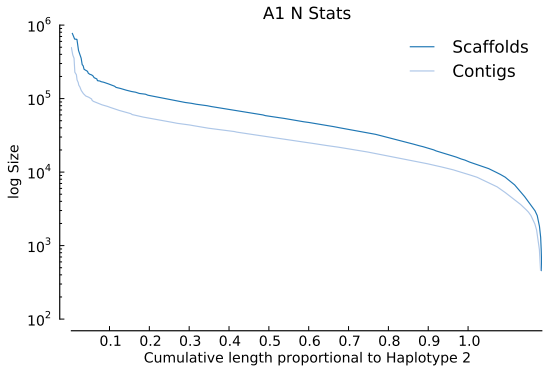
Assemblies:

A1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
C2	0.91842	0.91878	0.91805	0.00000
A1	0.90867	0.90879	0.90854	0.99971
I1	0.87175	0.87213	0.87138	0.99691

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	6,229	450	3,662.00	10,170	21,408.95	25,560.00	769,386	33,277.74	133,356,337
Contigs	9,741	450	3,122.00	8,008	13,654.16	18,021.00	494,128	17,752.25	133,005,137

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	89,278,573 – 89,999,824	80,139,373 – 80,644,836	160,276,348.0 – 161,282,878.0	1,199 – 3,397
Heterozygous	356,848 – 367,886	313,738 – 321,827	627,462.0 – 643,618.0	7 – 18
Indel	2,553,636 – 2,918,017	1,178,640 – 1,316,853	2,352,650.0 – 2,626,114.0	2,315 – 3,796

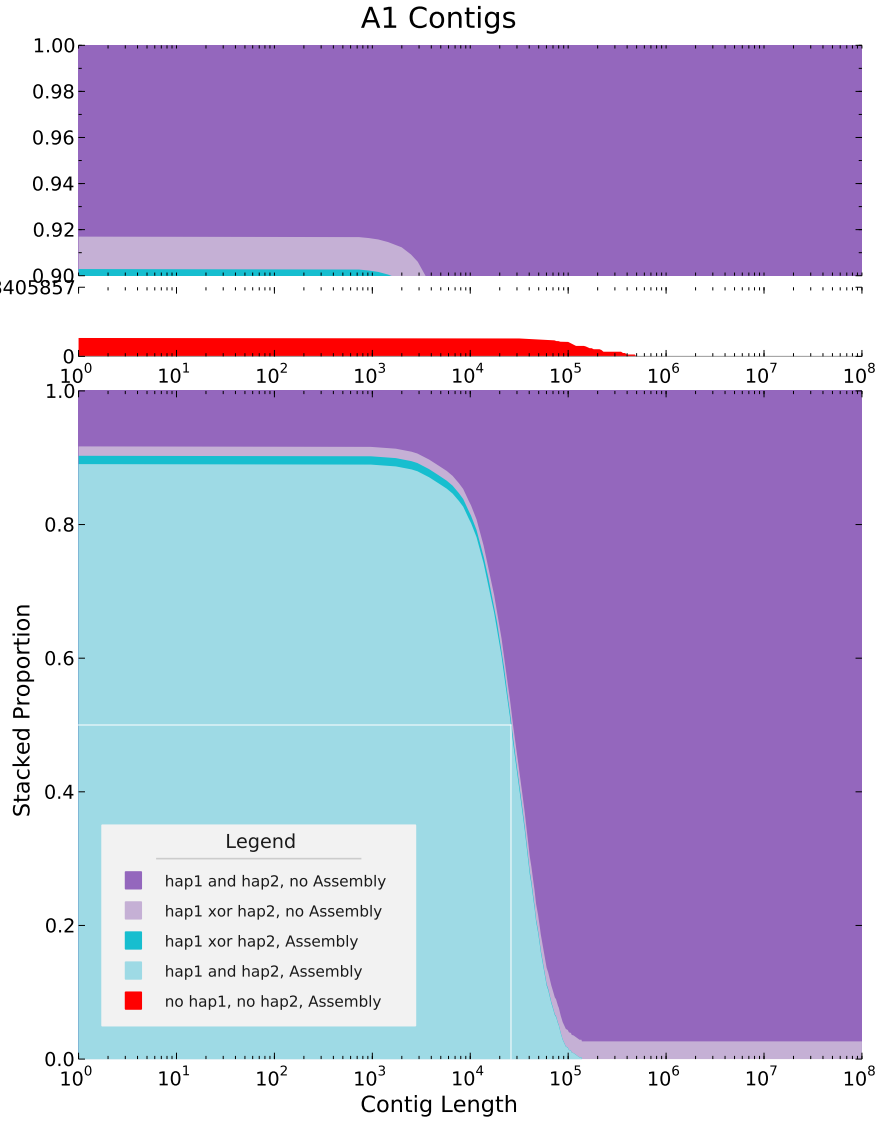


Figure 3.19: A1 contigs caption goes here.

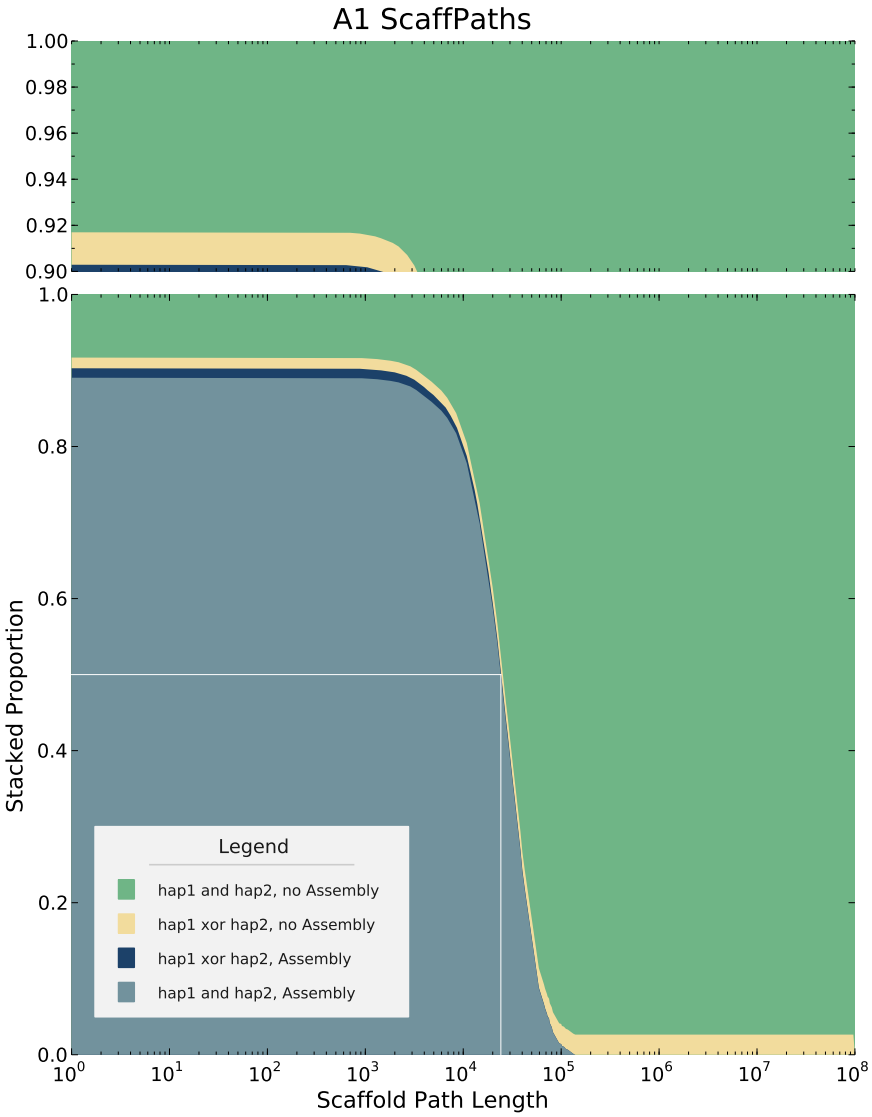


Figure 3.20: A1 scaffolds caption goes here.

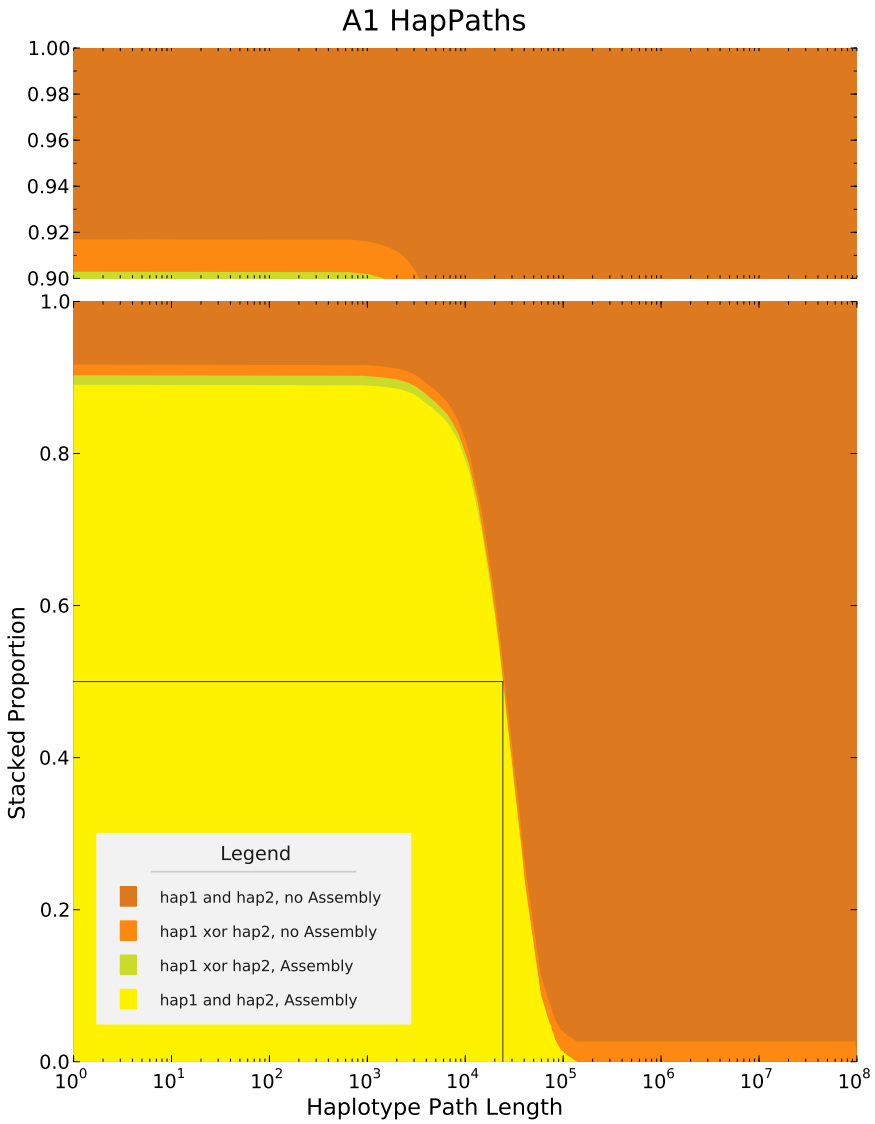


Figure 3.21: A1 hapPaths caption goes here.

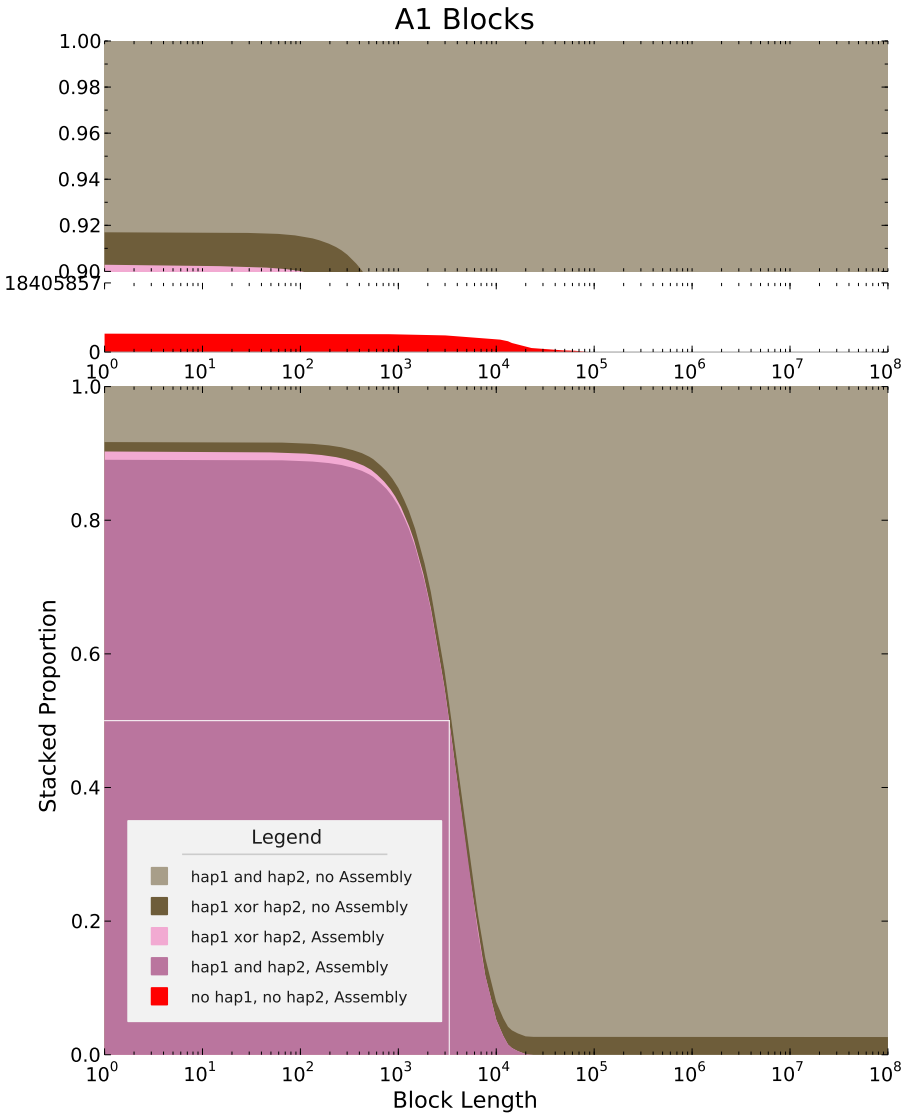


Figure 3.22: A1 blocks caption goes here.

3.2.2 B, Phusion

Affiliation: Wellcome Trust Sanger Institute, UK
Contact: Zemin Ning
Software: **Phusion2, phrap**
Number of entries: 2

ID	Total	Hap 1	Hap 2	Bac
B1	0.98694	0.98719	0.98668	0.99790
B2	0.98568	0.98600	0.98535	0.99892

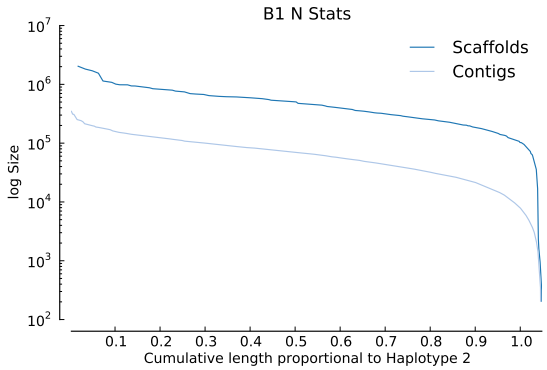
Assemblies:

B1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
P1	0.98852	0.98881	0.98823	0.00000
B1	0.98694	0.98719	0.98668	0.99790
F5	0.98691	0.98727	0.98653	0.99934

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	1,449	201	601.00	980	81,380.90	34,337.00	2,032,281	196,333.79	117,920,931
Contigs	4,502	201	2,057.25	10,489	26,125.20	35,388.00	347,708	37,024.75	117,615,631

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,960,814 – 109,581,319	108,388,158 – 108,973,477	216,773,514.0 – 217,940,394.0	1,401 – 3,280
Heterozygous	428,748 – 437,625	426,068 – 434,666	852,110.0 – 869,276.0	13 – 28
Indel	3,053,841 – 3,466,130	1,440,451 – 1,749,429	2,875,154.0 – 3,489,322.0	2,874 – 4,768

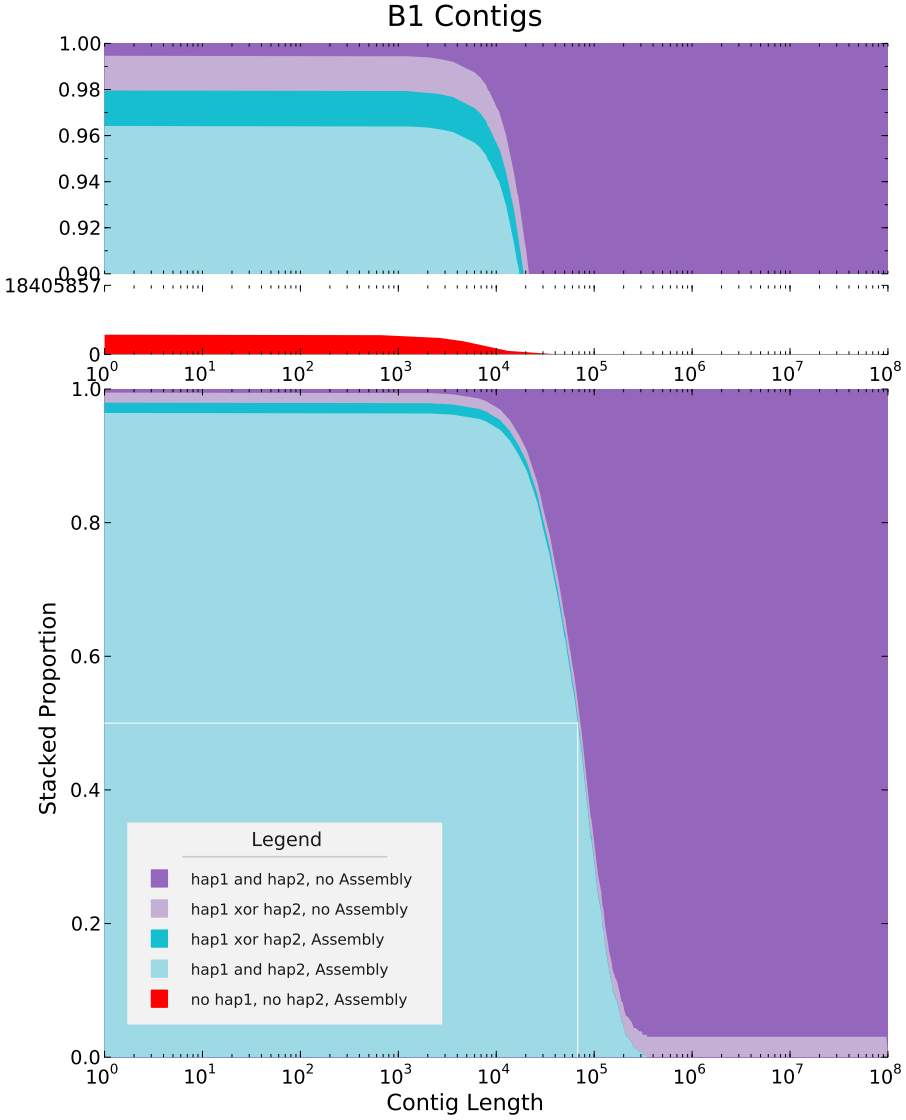


Figure 3.23: B1 contigs caption goes here.

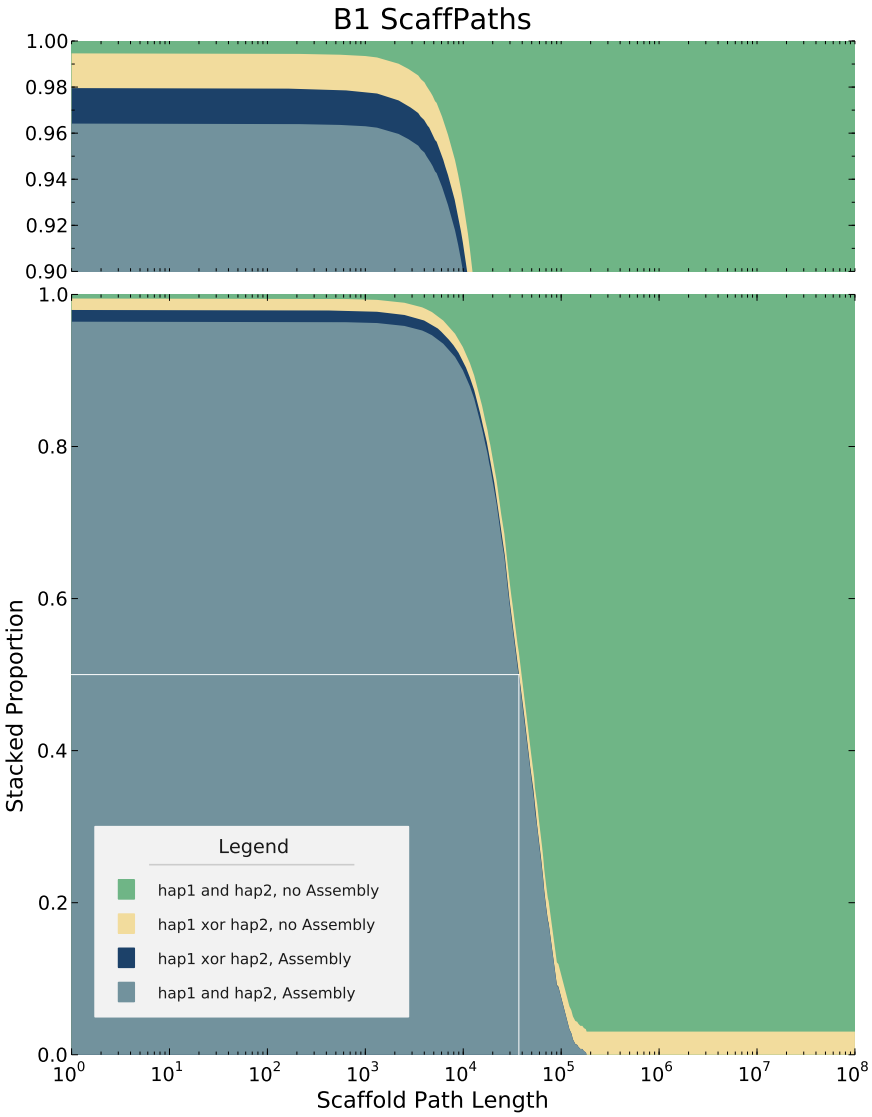


Figure 3.24: B1 scaffolds caption goes here.

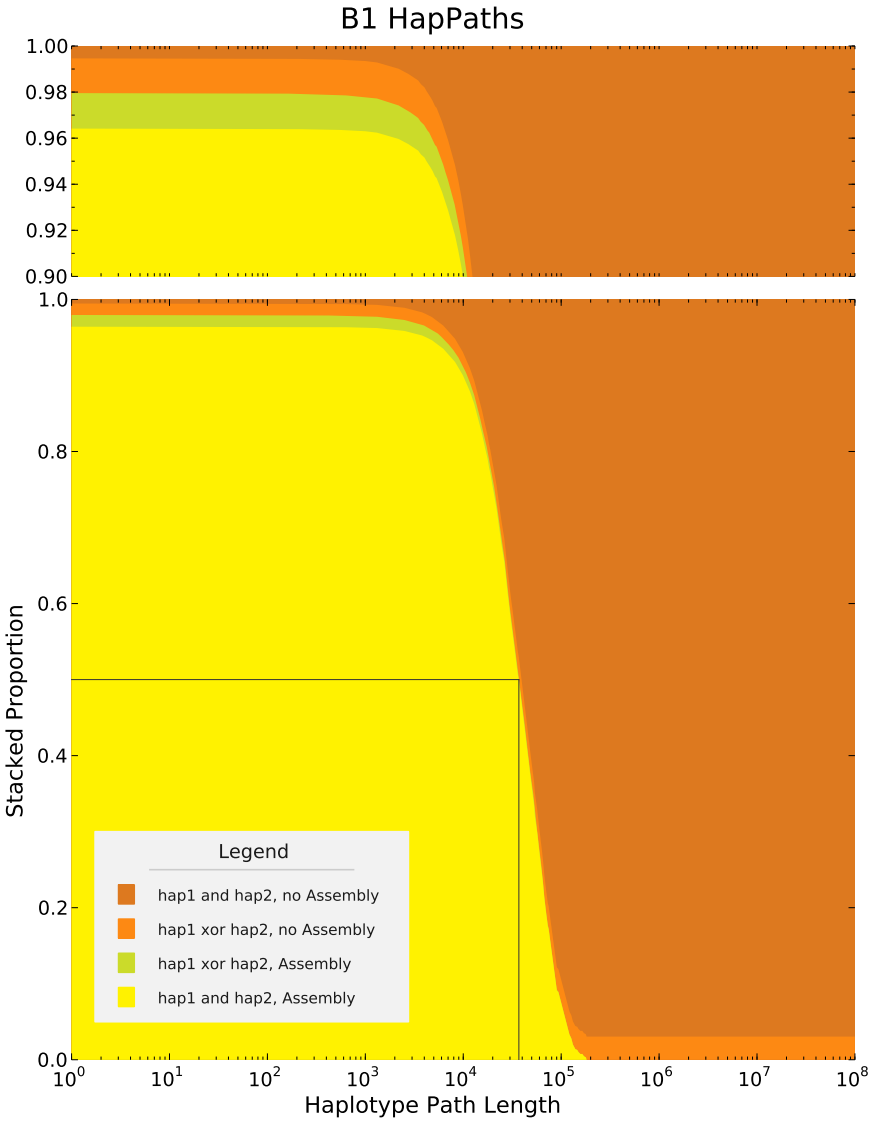


Figure 3.25: B1 hapPaths caption goes here.

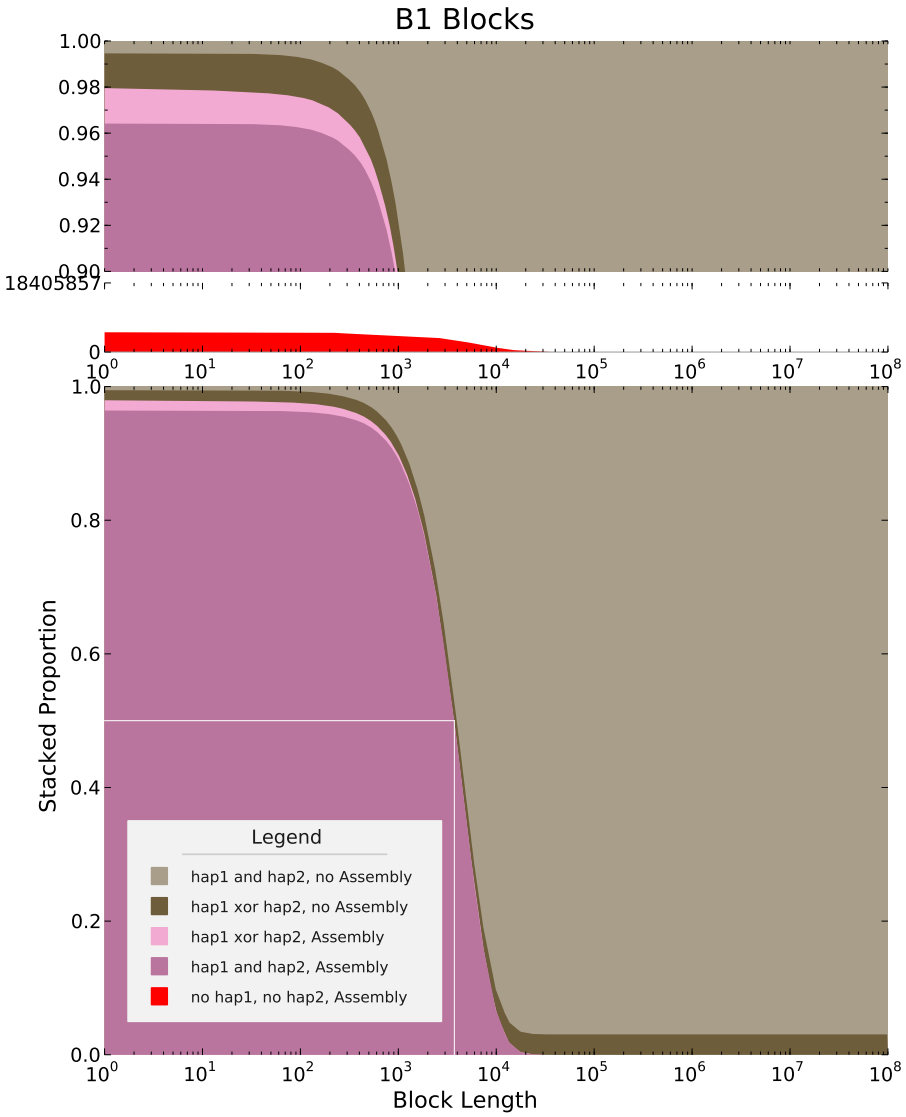


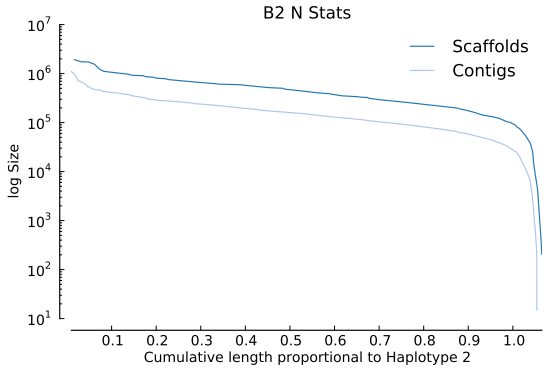
Figure 3.26: B1 blocks caption goes here.

B2

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
F1	0.98630	0.98664	0.98595	0.99923
B2	0.98568	0.98600	0.98535	0.99892
M2	0.98534	0.98552	0.98516	0.99969

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	2,048	200	344.75	730	58,515.75	5,858.25	1,922,525	166,675.08	119,840,251
Contigs	3,040	15	420.75	1,455	39,007.87	47,076.50	1,110,065	78,927.82	118,583,921

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,028,121 – 108,761,484	107,315,720 – 107,958,959	214,624,640.0 – 215,903,846.0	3,299 – 6,754
Heterozygous	419,996 – 433,136	416,349 – 428,442	832,670.0 – 856,792.0	12 – 35
Indel	2,732,421 – 3,188,445	1,280,179 – 1,594,481	2,555,814.0 – 3,181,080.0	2,259 – 3,894

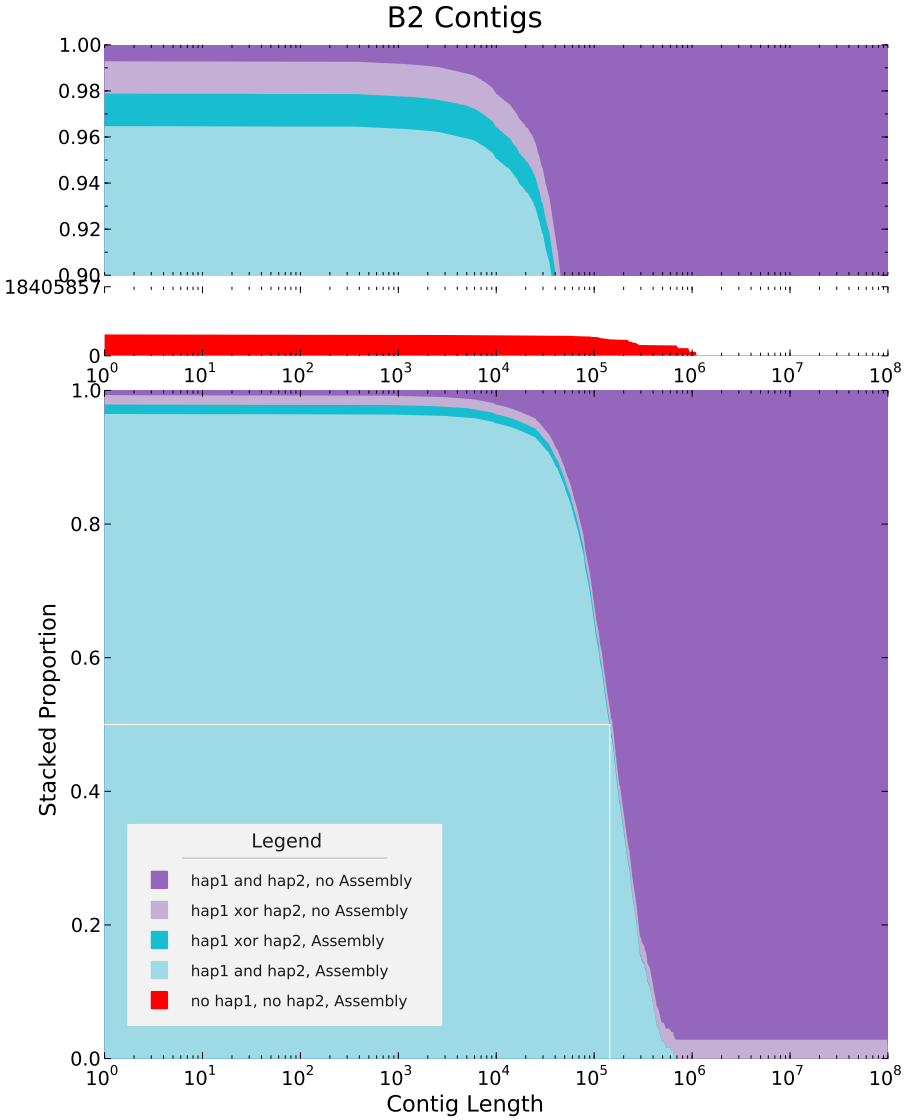


Figure 3.27: B2 contigs caption goes here.

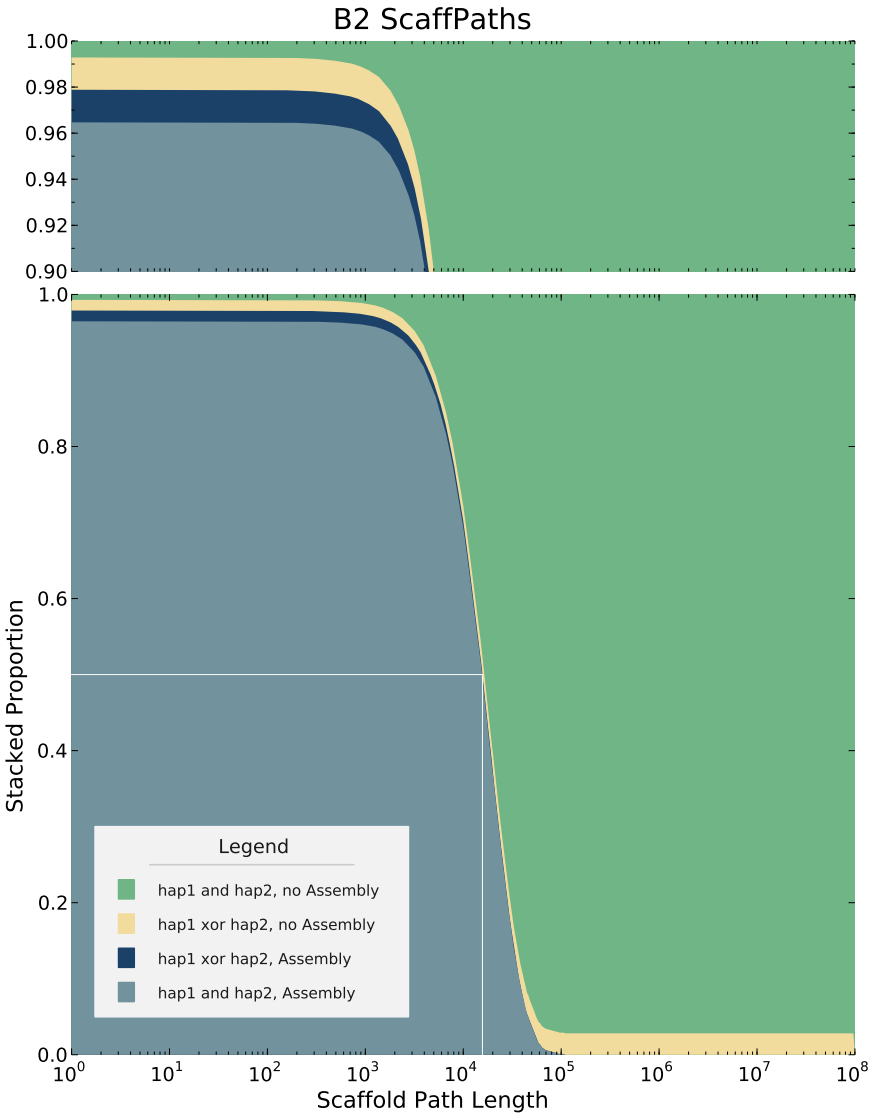


Figure 3.28: B2 scaffolds caption goes here.

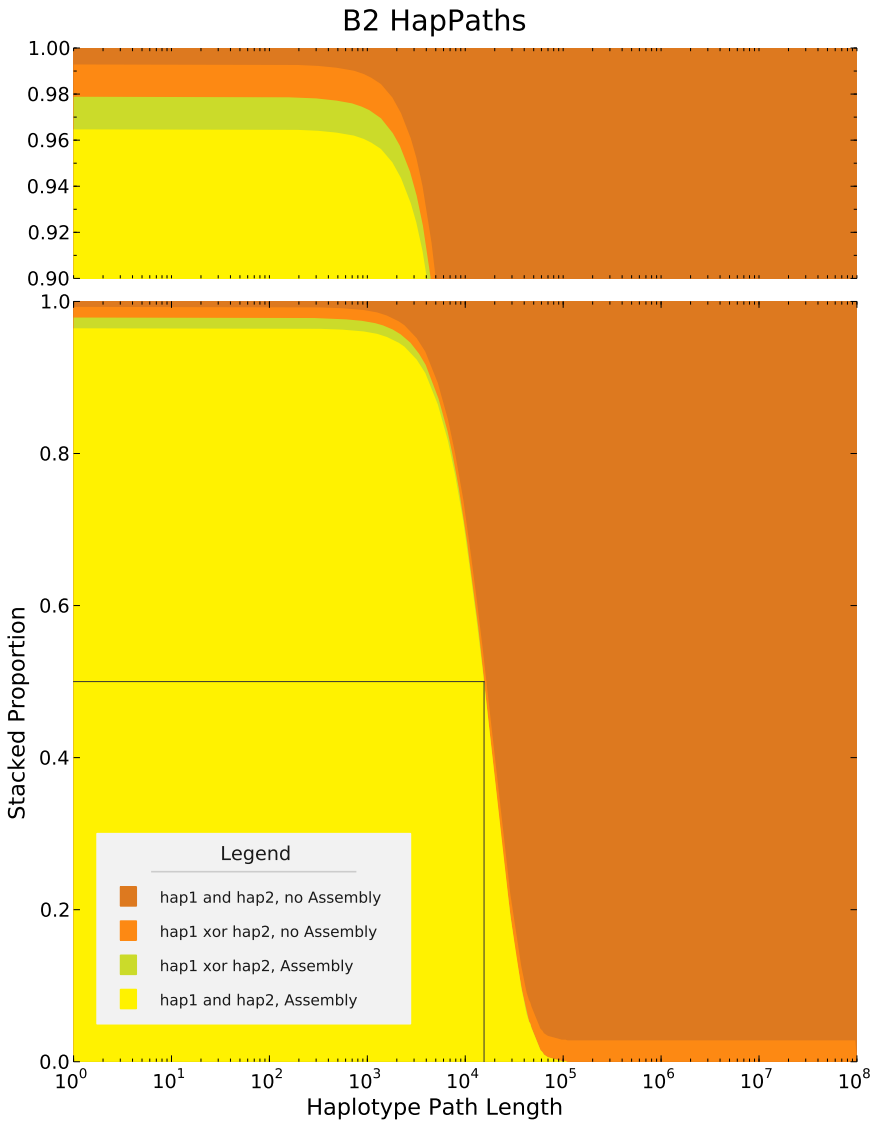


Figure 3.29: B2 hapPaths caption goes here.

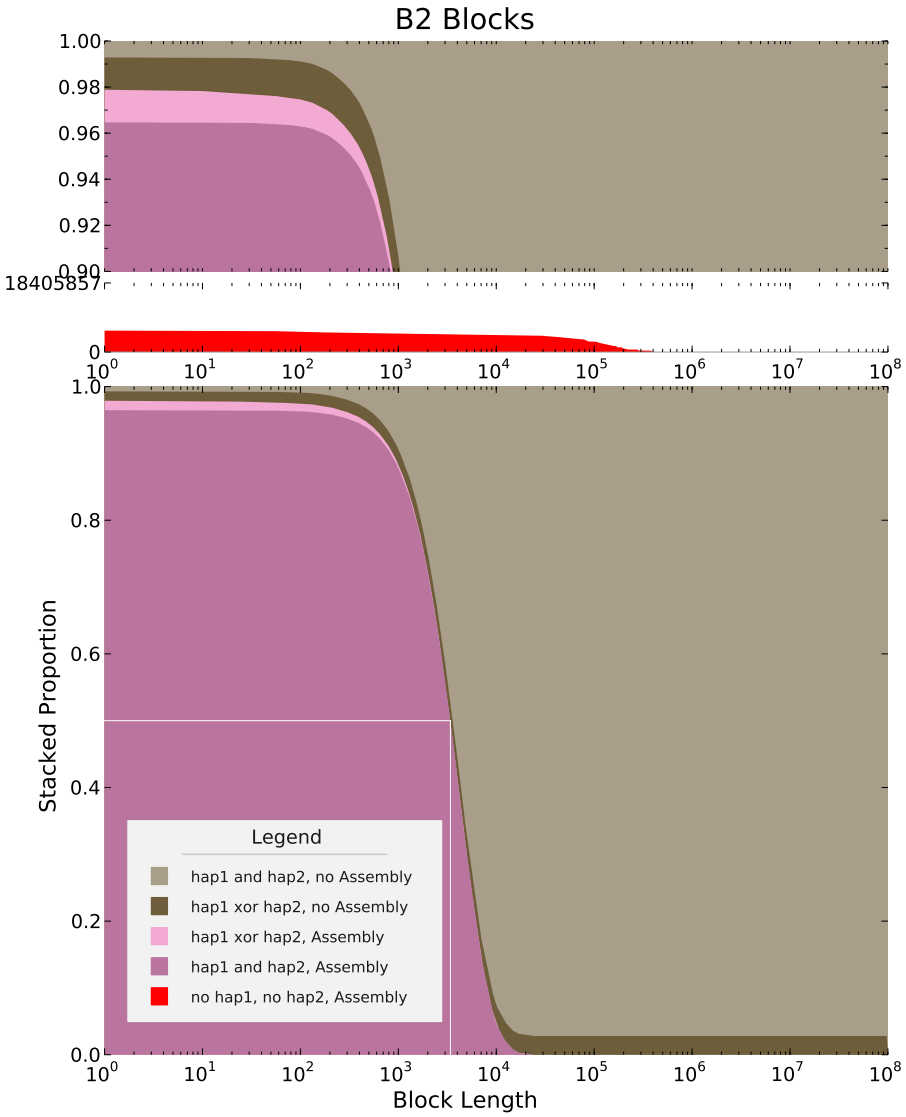


Figure 3.30: B2 blocks caption goes here.

3.2.3 C, Ensembl Genomes' Curtain

Affiliation: European Bioinformatics Institute, UK
Contact: Matthias Haimel
Software: **SGA, BWA, Curtain, Velvet**
Number of entries: 2

ID	Total	Hap 1	Hap 2	Bac
C1	0.97348	0.97347	0.97348	0.00903
C2	0.91842	0.91878	0.91805	0.00000

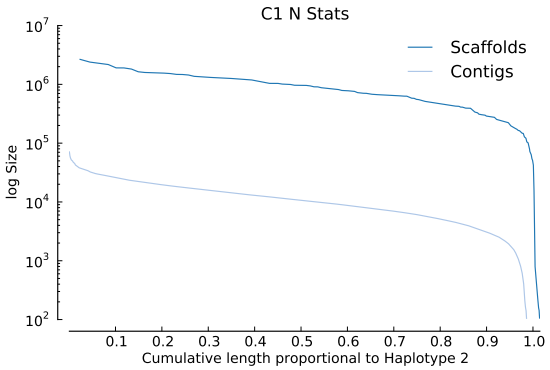
Assemblies:

C1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
X2	0.97492	0.97516	0.97467	0.99848
C1	0.97348	0.97347	0.97348	0.00903
X1	0.97305	0.97332	0.97277	0.99869

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	5,249	105	135.00	172	21,755.22	359.00	2,660,971	147,634.79	114,193,167
Contigs	20,229	105	668.00	3,318	5,488.52	7,858.00	70,936	6,488.19	111,027,307

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	107,995,683 – 109,484,691	106,928,115 – 107,548,512	213,851,312.0 – 215,077,024.0	615 – 900
Heterozygous	410,716 – 432,828	401,911 – 408,634	803,492.0 – 816,264.0	1 – 1
Indel	2,826,068 – 3,222,508	1,212,471 – 1,347,183	2,420,966.0 – 2,685,340.0	1,958 – 2,329

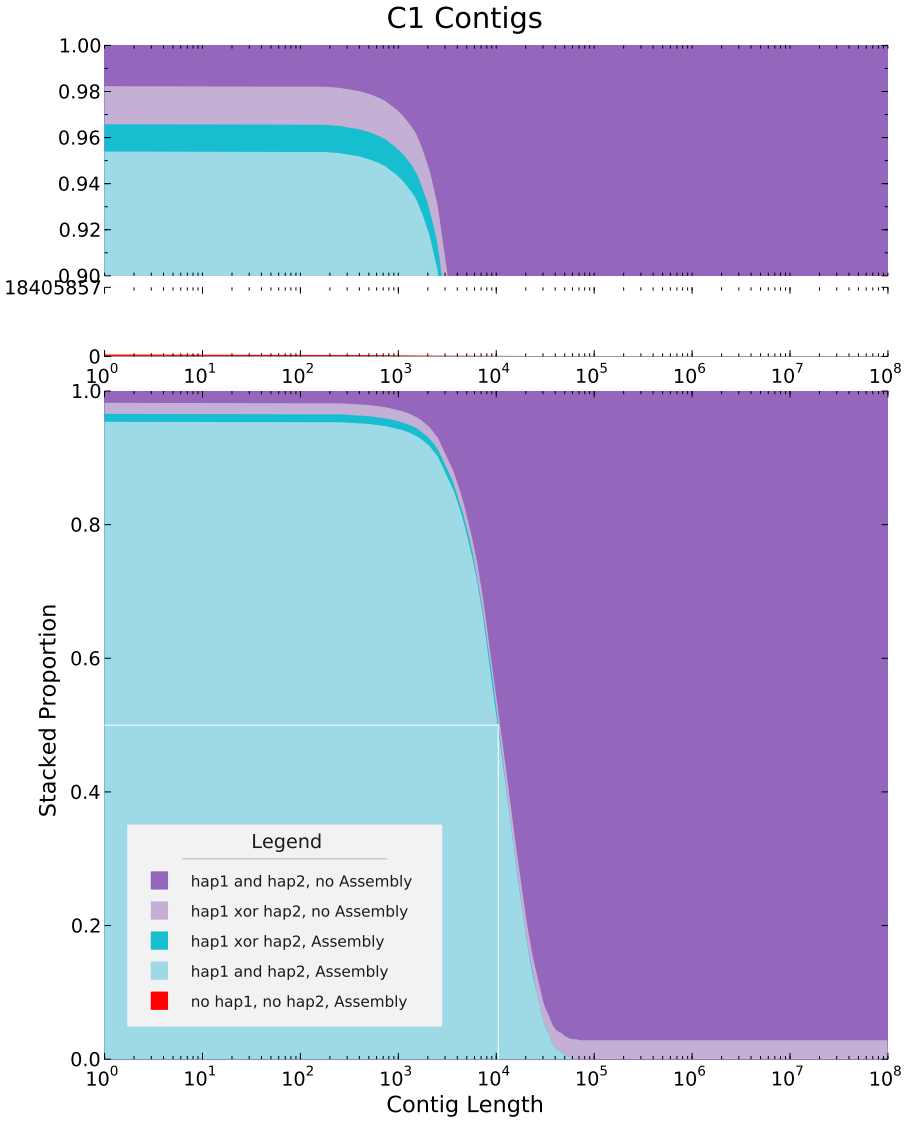


Figure 3.31: C1 contigs caption goes here.

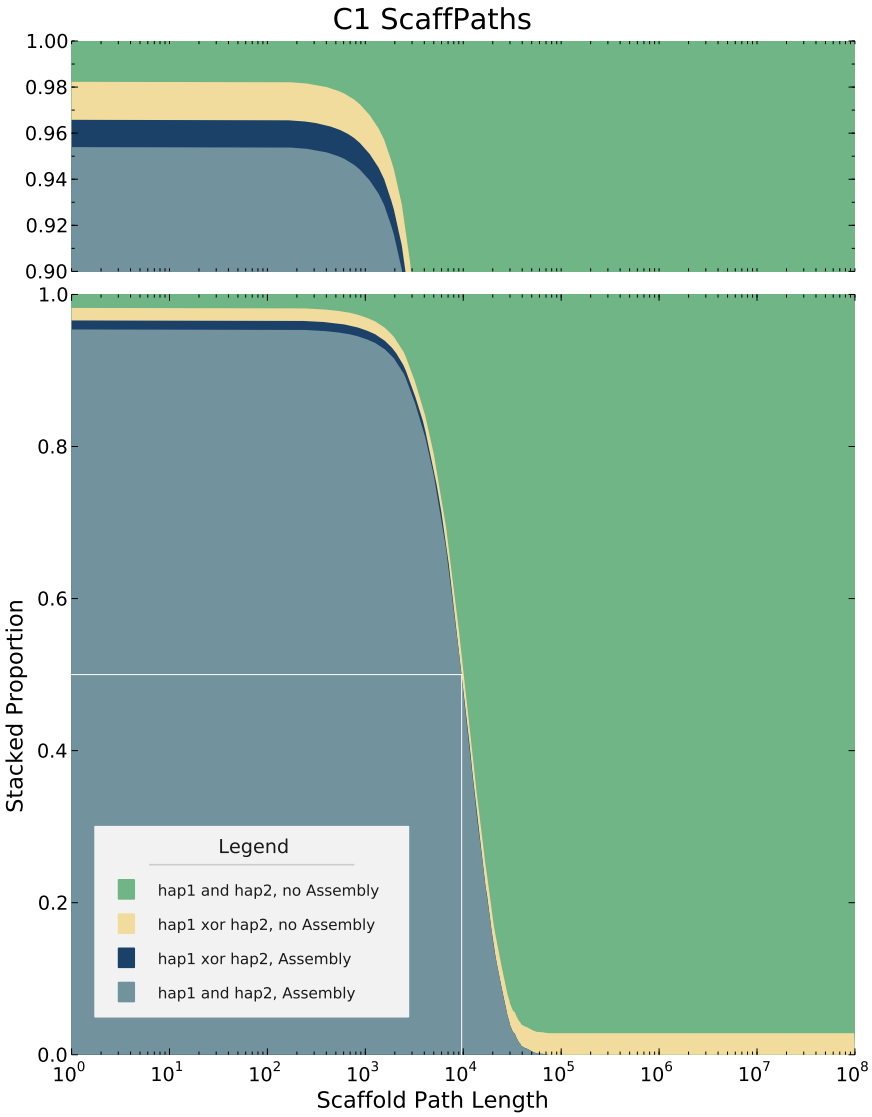


Figure 3.32: C1 scaffolds caption goes here.

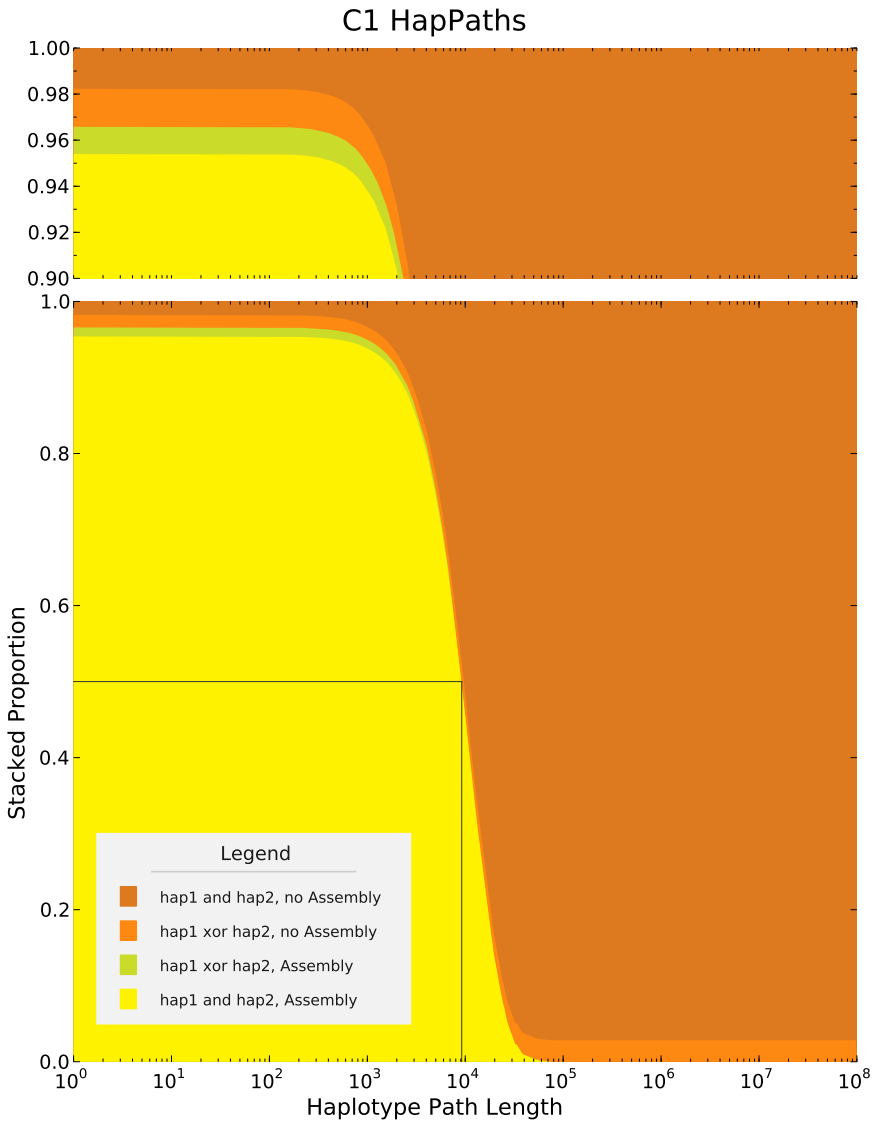


Figure 3.33: C1 hapPaths caption goes here.

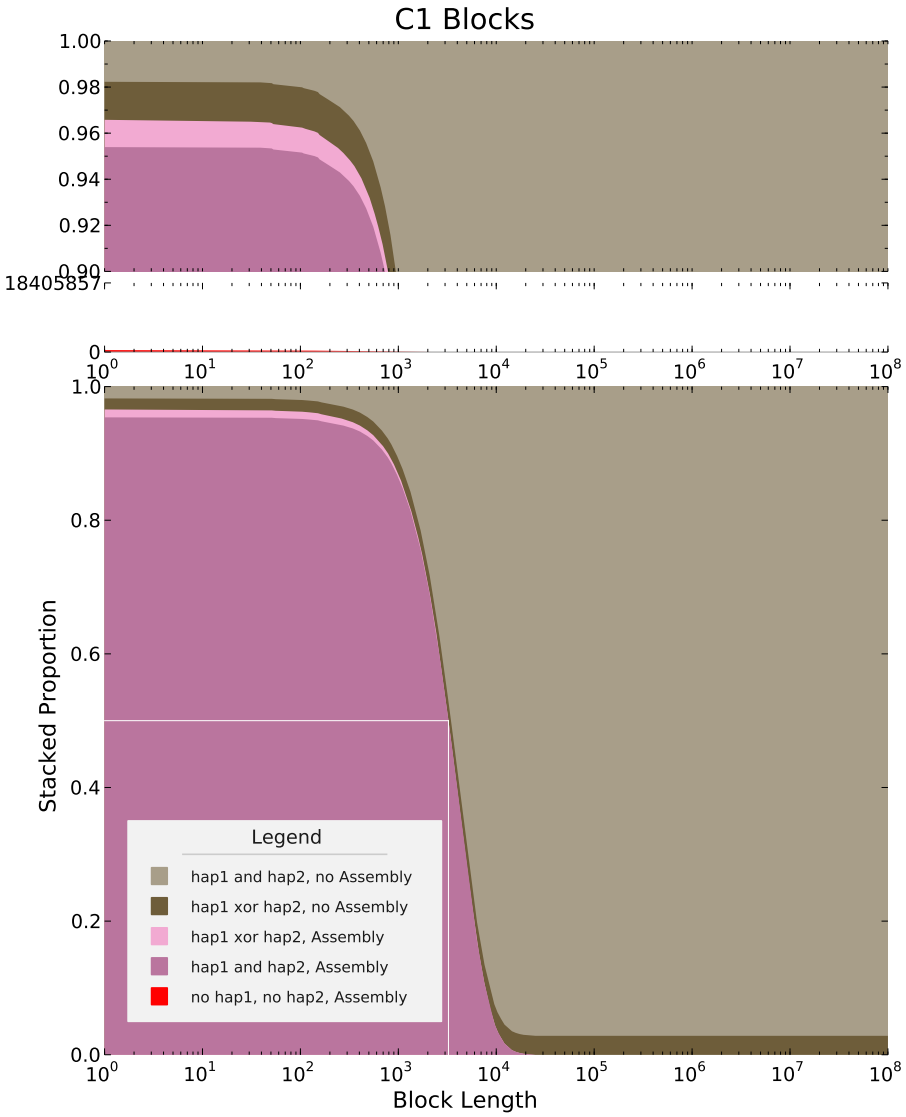


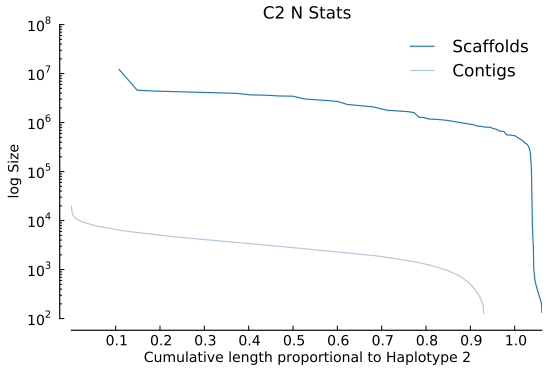
Figure 3.34: C1 blocks caption goes here.

C2

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
H2	0.92711	0.92728	0.92695	0.99536
C2	0.91842	0.91878	0.91805	0.00000
A1	0.90867	0.90879	0.90854	0.99971

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	6,114	135	240.00	309	19,530.76	426.00	12,125,519	261,980.80	119,411,083
Contigs	53,273	126	626.00	1,500	1,964.32	2,700.00	19,554	1,758.50	104,645,412

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	107,032,299 – 110,591,207	101,422,736 – 102,324,638	202,817,188.0 – 204,485,796.0	64 – 702
Heterozygous	377,060 – 441,719	323,830 – 332,130	643,768.0 – 650,916.0	2 – 6
Indel	1,506,313 – 1,874,511	549,592 – 593,180	1,097,682.0 – 1,165,892.0	676 – 817

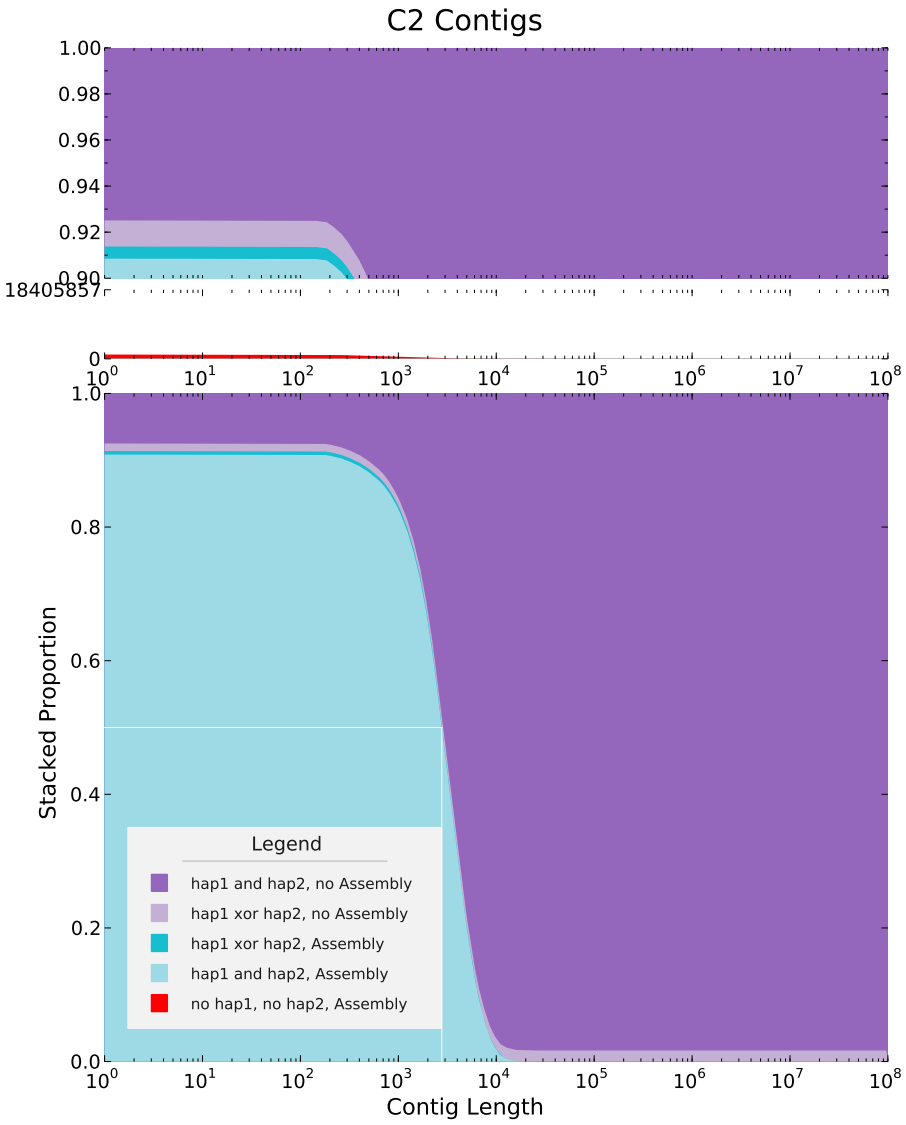


Figure 3.35: C2 contigs caption goes here.

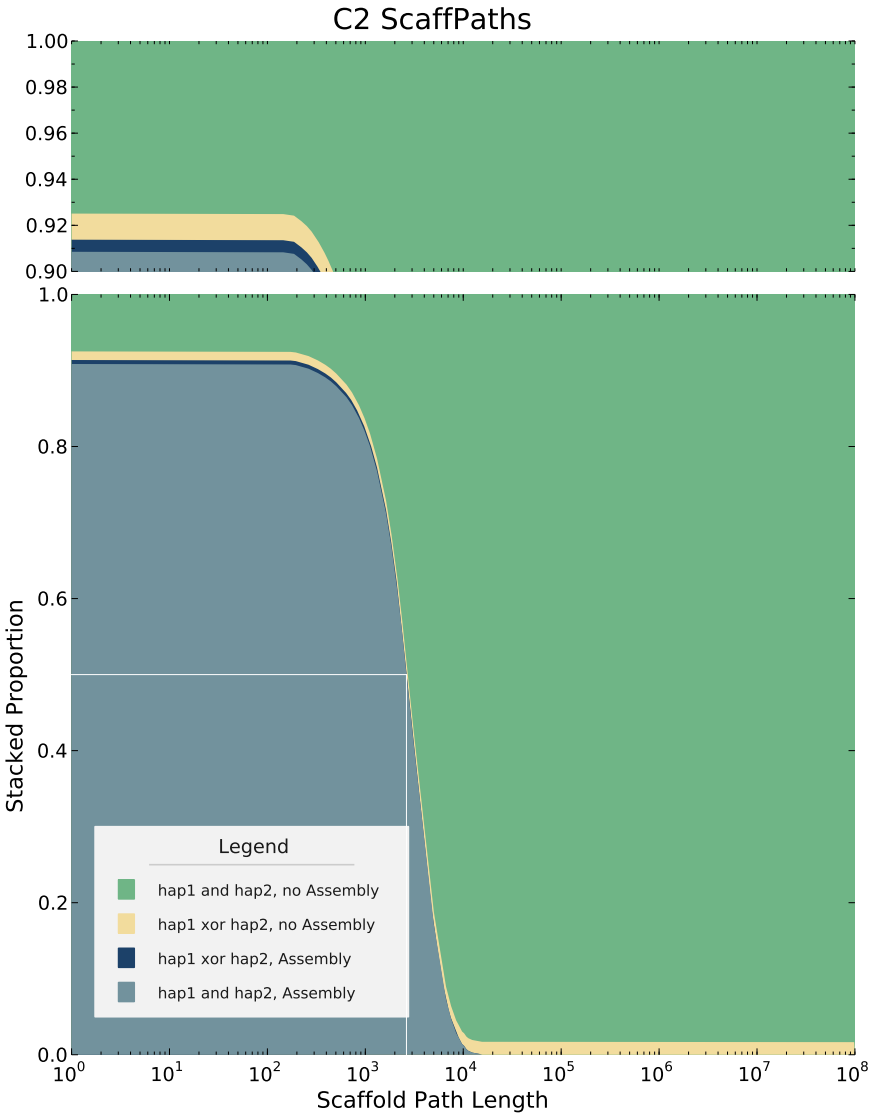


Figure 3.36: C2 scaffolds caption goes here.

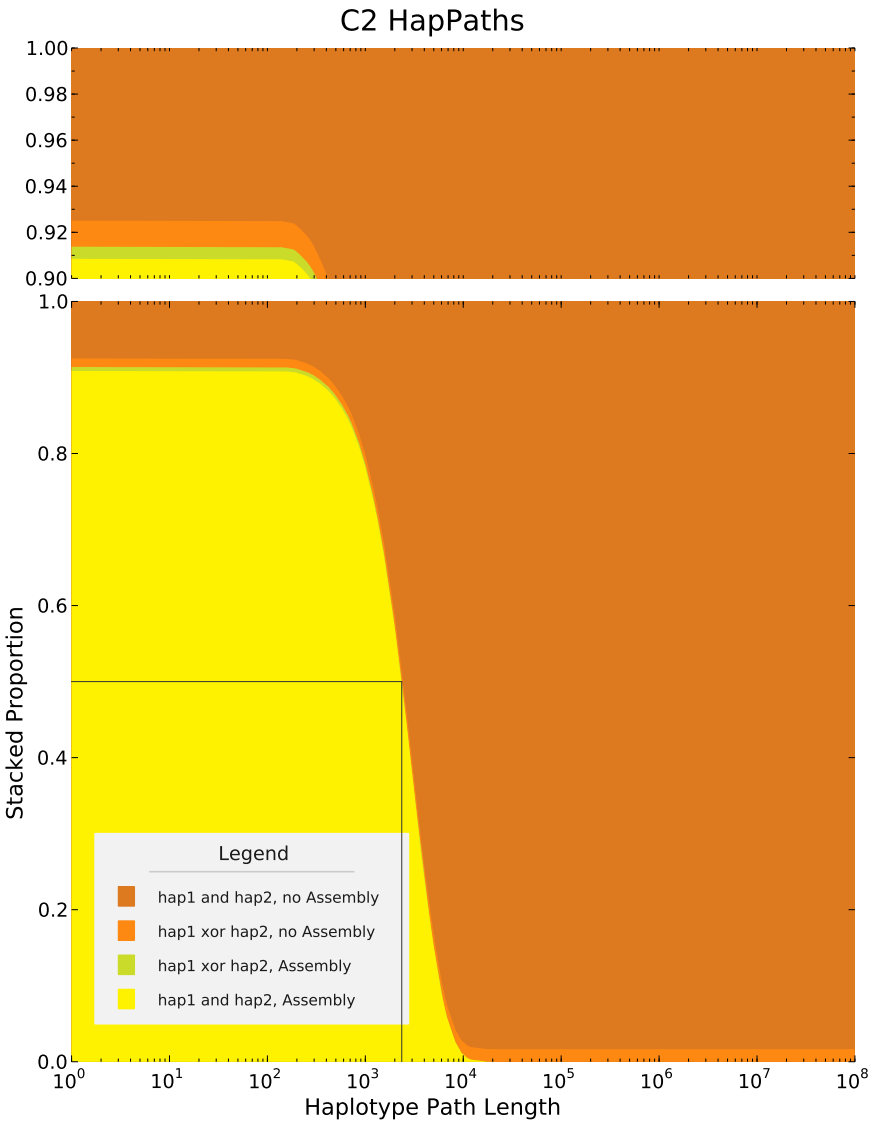


Figure 3.37: C2 hapPaths caption goes here.

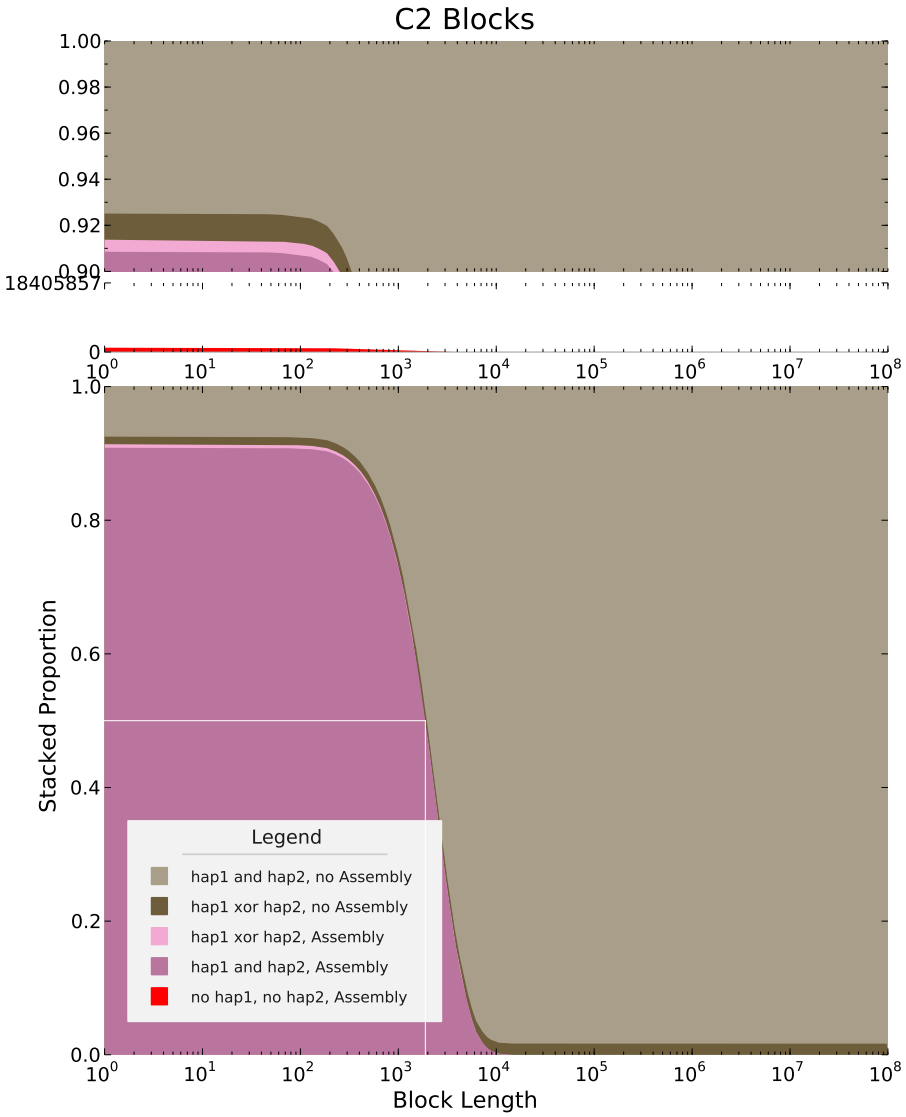


Figure 3.38: C2 blocks caption goes here.

3.2.4 D, sanger-sga

Affiliation: Wellcome Trust Sanger Insitute, UK
Contact: Jared Simpson
Software: **SGA**
Number of entries: 4

ID	Total	Hap 1	Hap 2	Bac
D4	0.98288	0.98303	0.98274	0.99618
D2	0.97705	0.97715	0.97696	0.99086
D1	0.97616	0.97642	0.97591	0.99086
D3	0.97616	0.97634	0.97595	0.99085

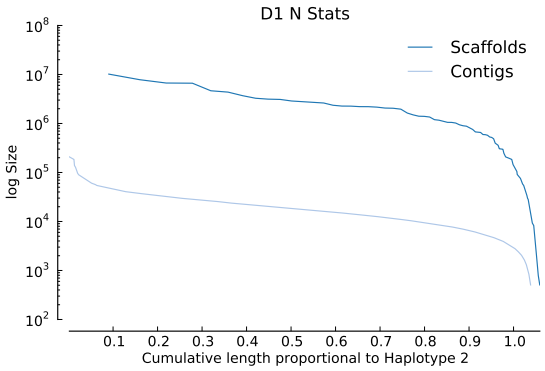
Assemblies:

D1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
D2	0.97705	0.97715	0.97696	0.99086
D1	0.97616	0.97642	0.97591	0.99086
D3	0.97616	0.97634	0.97595	0.99085

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	1,433	500	654.00	860	83,164.91	3,046.00	10,172,515	546,109.42	119,175,323
Contigs	11,067	500	2,847.00	7,012	10,559.40	14,365.00	207,957	11,989.39	116,860,825

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	107,772,915 – 108,544,819	106,406,177 – 106,922,341	212,812,342.0 – 213,844,660.0	6 – 11
Heterozygous	419,027 – 433,432	411,097 – 417,224	822,194.0 – 834,448.0	0 – 0
Indel	2,705,290 – 3,077,308	1,167,503 – 1,269,482	2,331,934.0 – 2,535,748.0	1,536 – 1,608

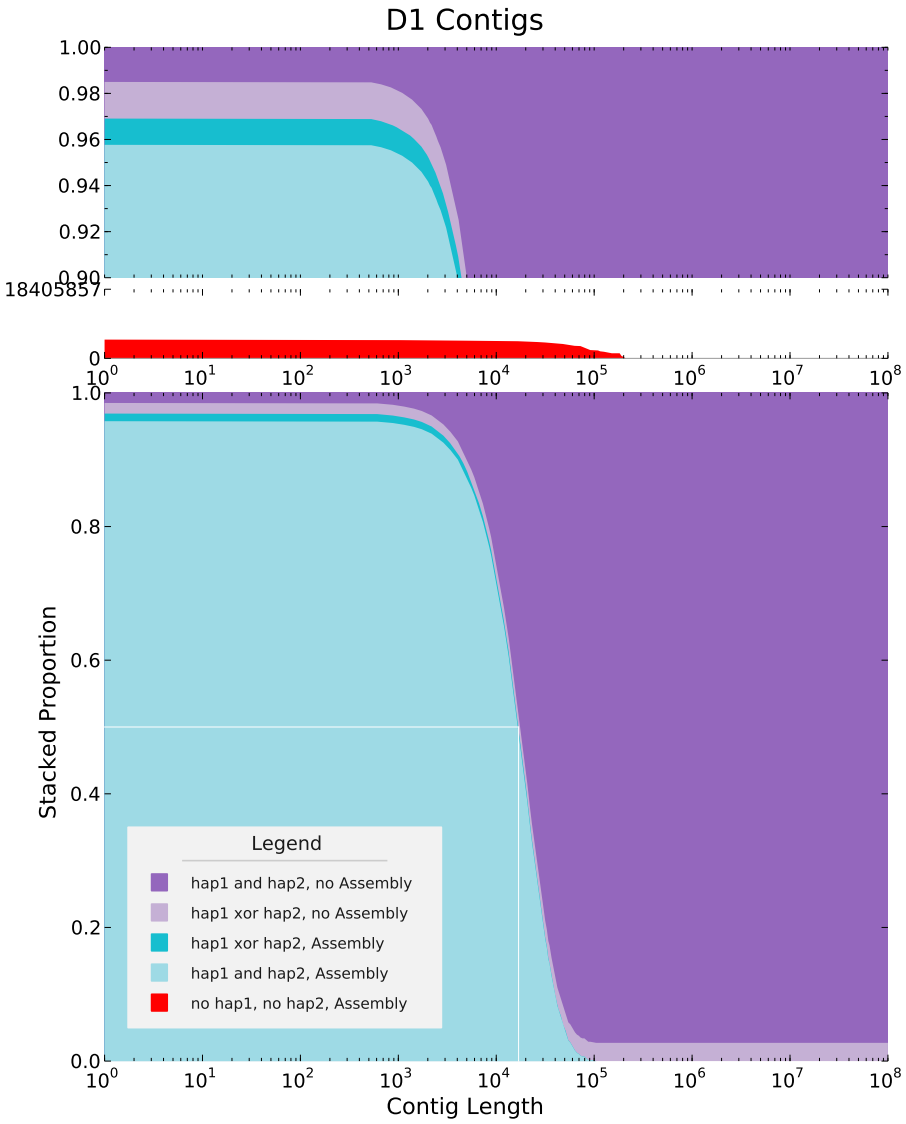


Figure 3.39: D1 contigs caption goes here.

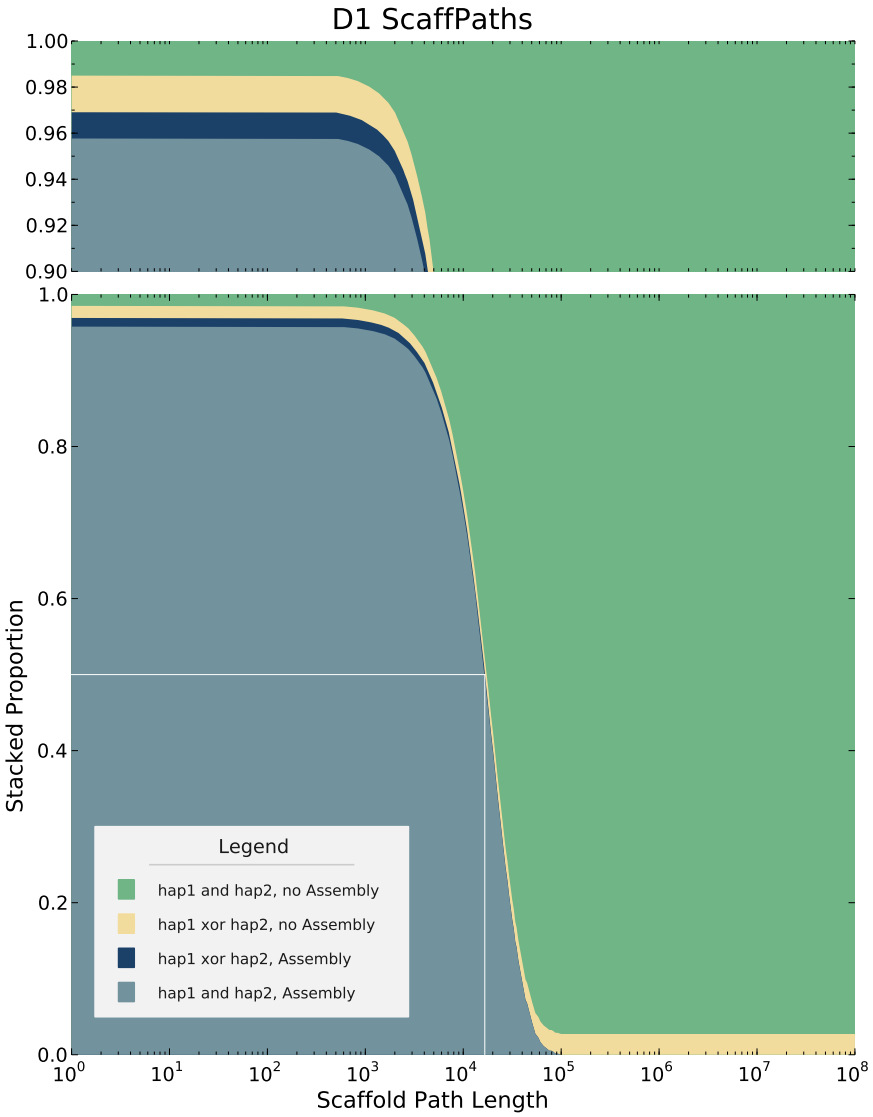


Figure 3.40: D1 scaffolds caption goes here.

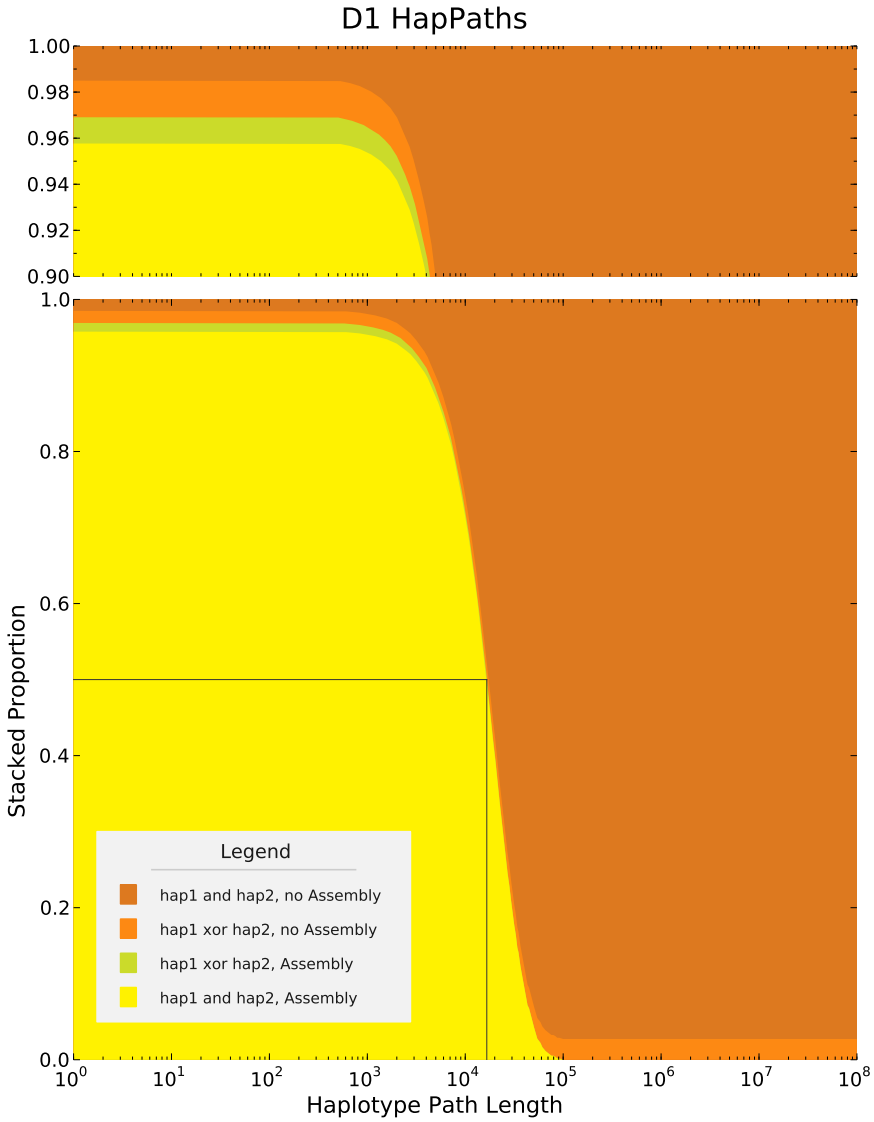


Figure 3.41: D1 hapPaths caption goes here.

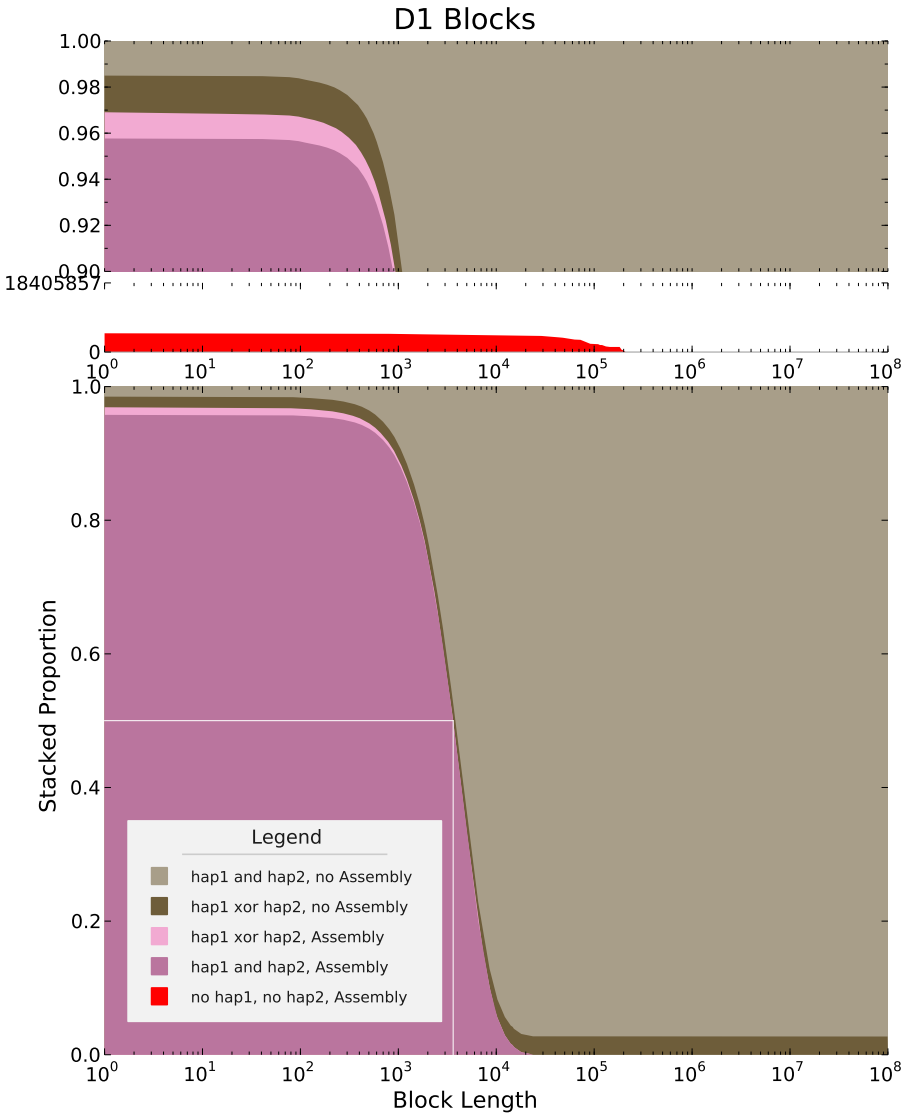


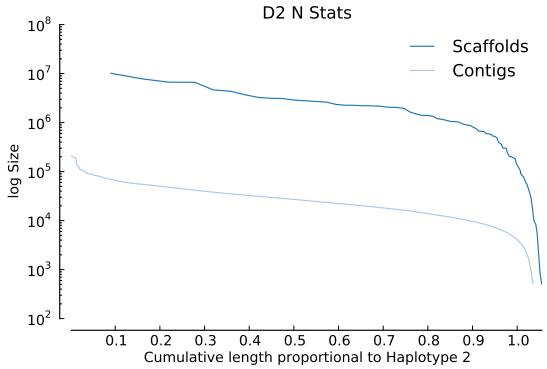
Figure 3.42: D1 blocks caption goes here.

D2

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
K3	0.98038	0.98042	0.98034	0.00017
D2	0.97705	0.97715	0.97696	0.99086
D1	0.97616	0.97642	0.97591	0.99086

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	1,381	500	652.00	854	85,981.21	2,770.00	10,167,372	555,833.03	118,740,053
Contigs	7,904	500	3,229.00	9,616	14,745.53	20,287.00	207,957	16,703.87	116,548,634

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,141,752 – 108,865,817	106,846,982 – 107,352,006	213,693,938.0 – 214,703,976.0	13 – 18
Heterozygous	421,157 – 434,767	413,511 – 419,945	827,022.0 – 839,890.0	0 – 0
Indel	2,772,129 – 3,143,607	1,213,099 – 1,314,510	2,422,842.0 – 2,625,564.0	1,678 – 1,728

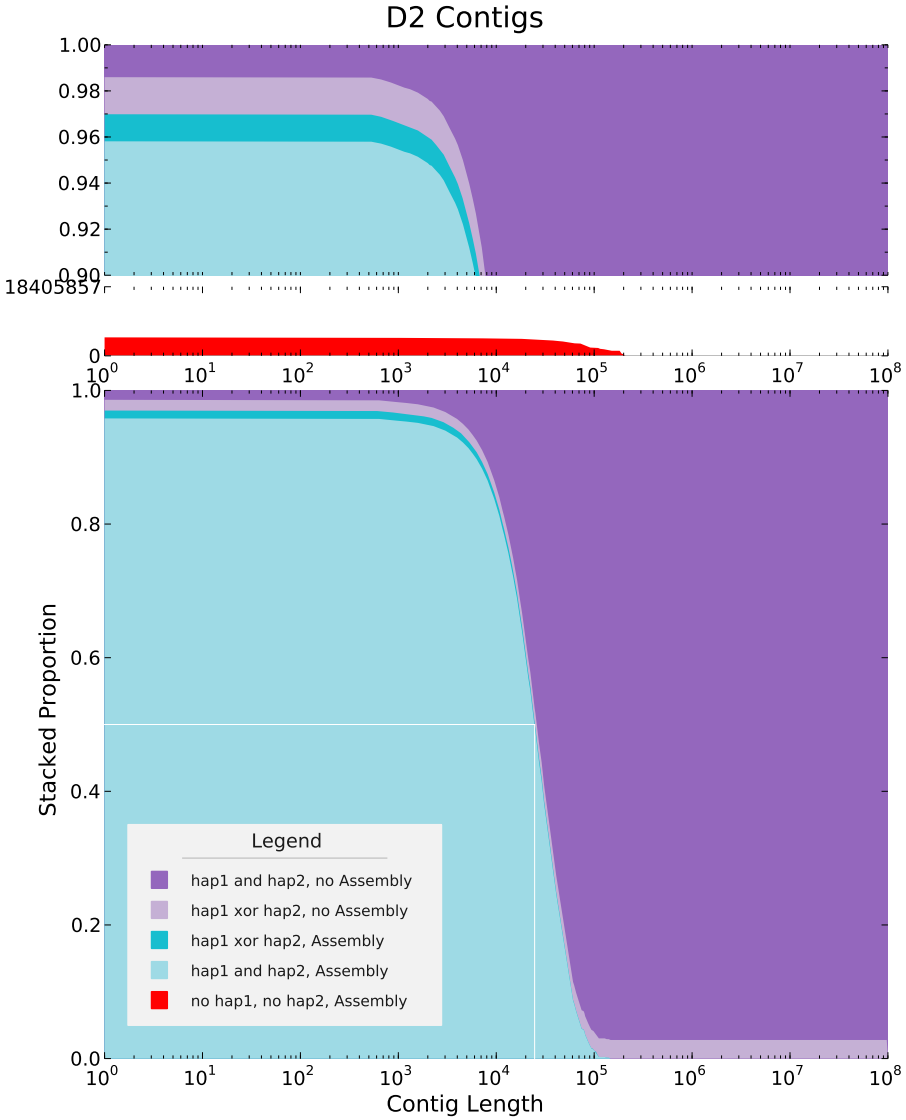


Figure 3.43: D2 contigs caption goes here.

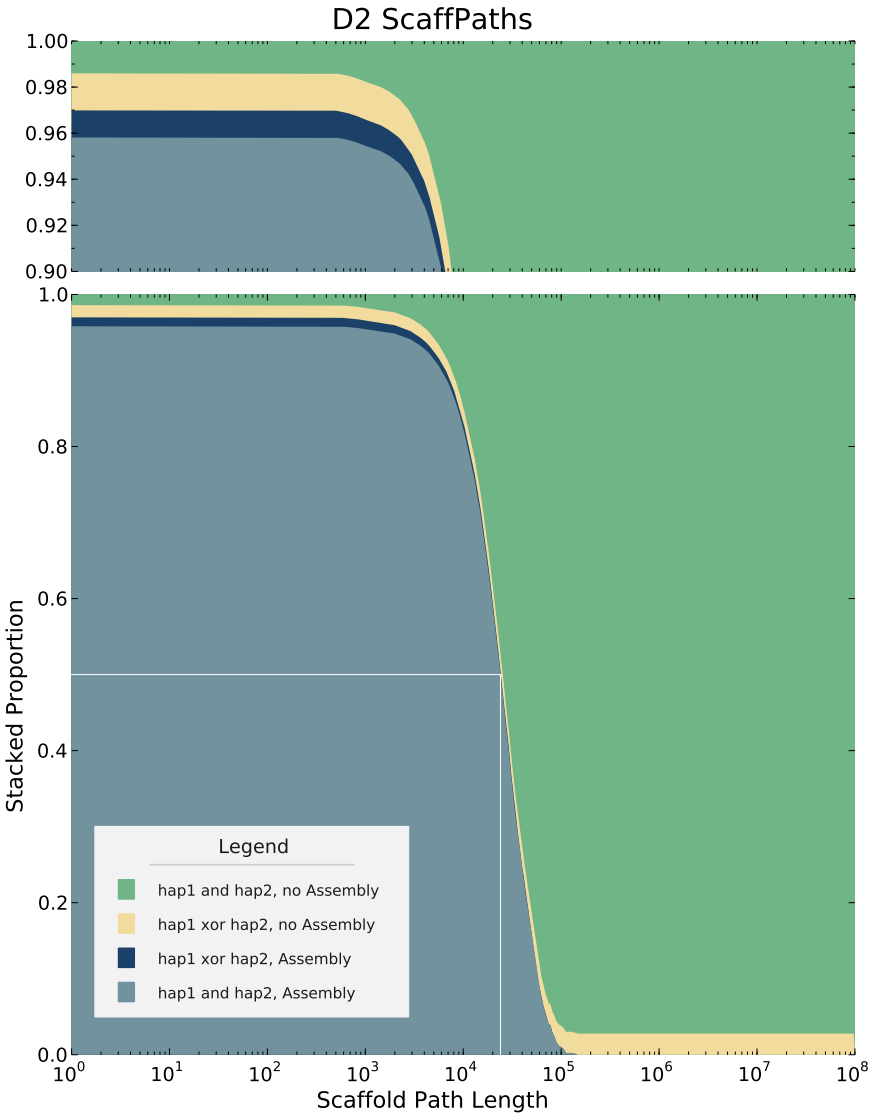


Figure 3.44: D2 scaffolds caption goes here.

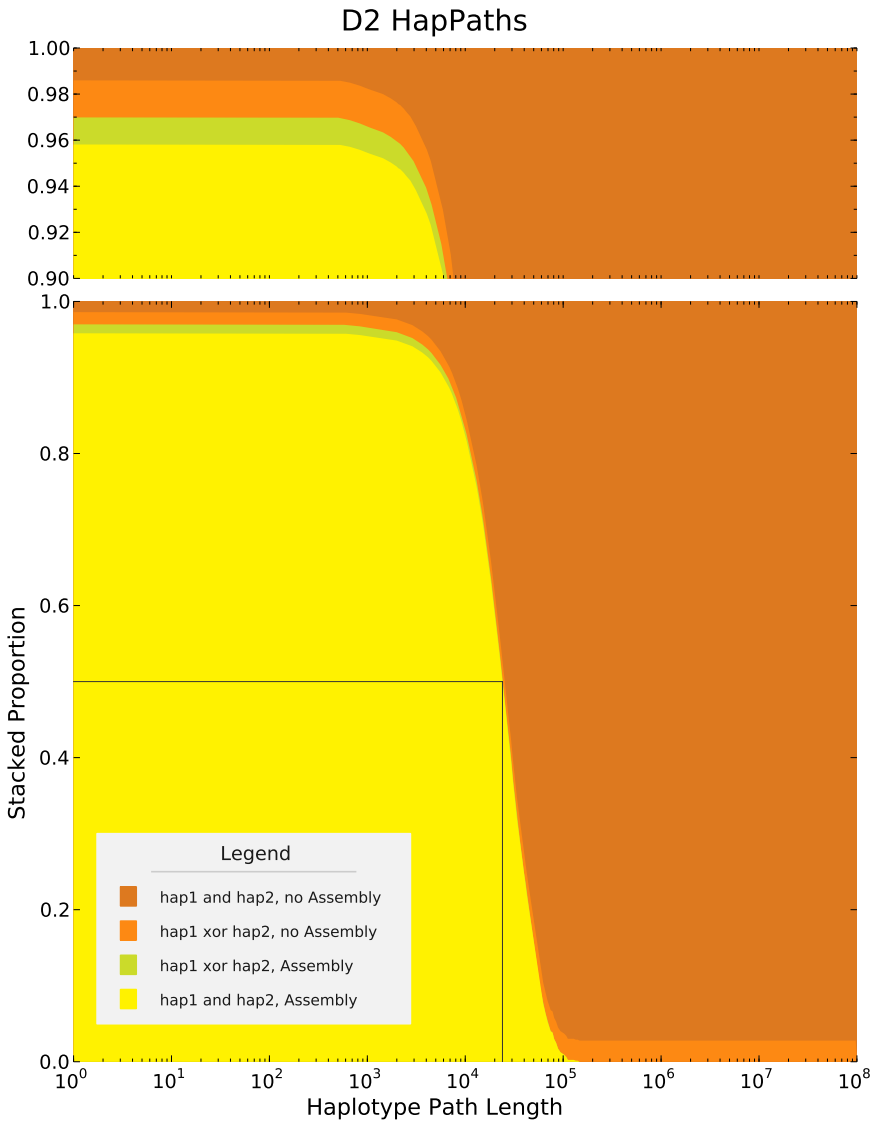


Figure 3.45: D2 hapPaths caption goes here.

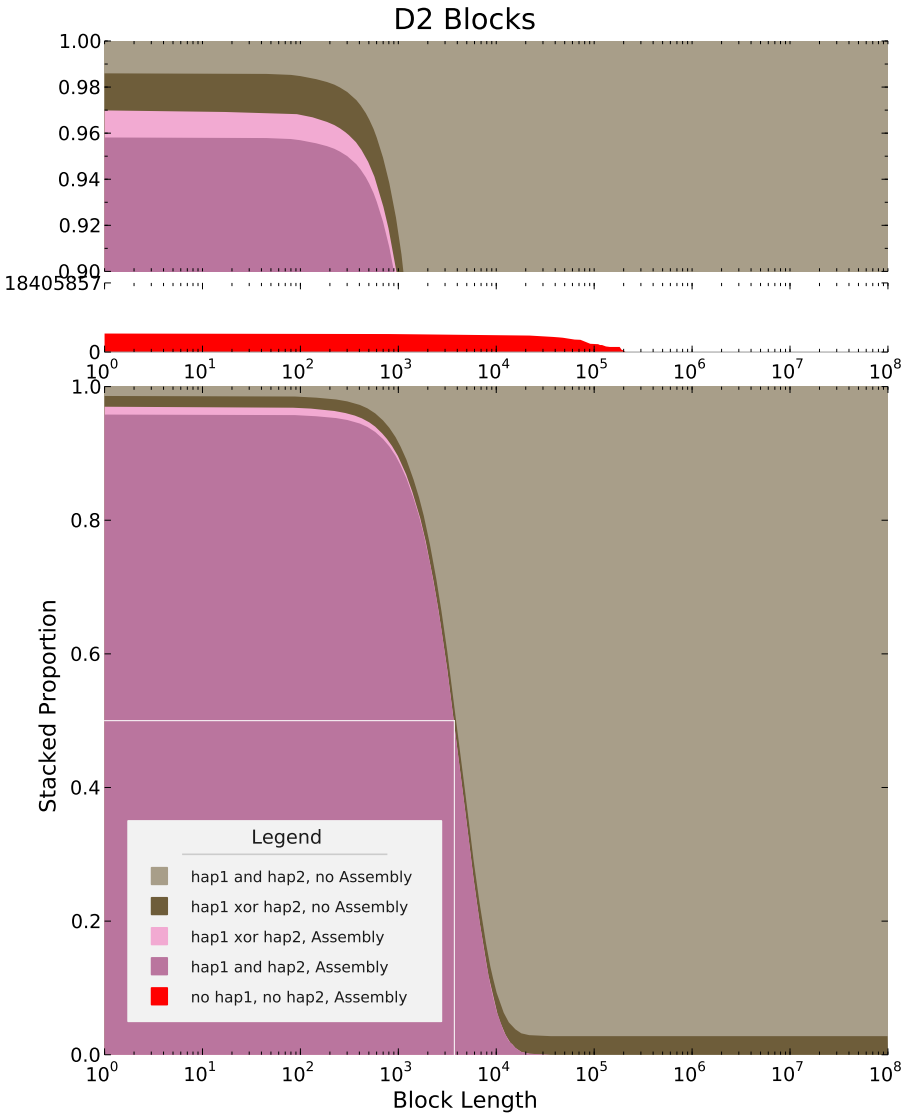


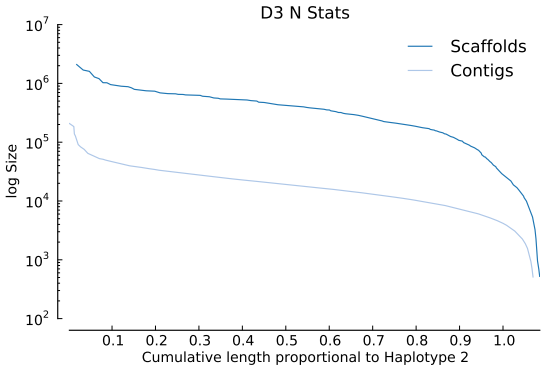
Figure 3.46: D2 blocks caption goes here.

D3

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
D1	0.97616	0.97642	0.97591	0.99086
D3	0.97616	0.97634	0.97595	0.99085
X2	0.97492	0.97516	0.97467	0.99848

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	2,429	500	792.00	3,132	50,253.92	18,467.00	2,094,112	144,852.59	122,066,780
Contigs	11,310	500	2,888.00	7,077	10,639.09	14,507.25	207,957	11,979.60	120,328,105

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	104,350,290 – 105,105,307	102,997,160 – 103,498,102	205,994,308.0 – 206,996,180.0	6 – 12
Heterozygous	405,710 – 419,981	397,846 – 403,850	795,692.0 – 807,700.0	0 – 0
Indel	2,733,071 – 3,103,547	1,168,770 – 1,269,450	2,334,366.0 – 2,535,592.0	1,587 – 1,654

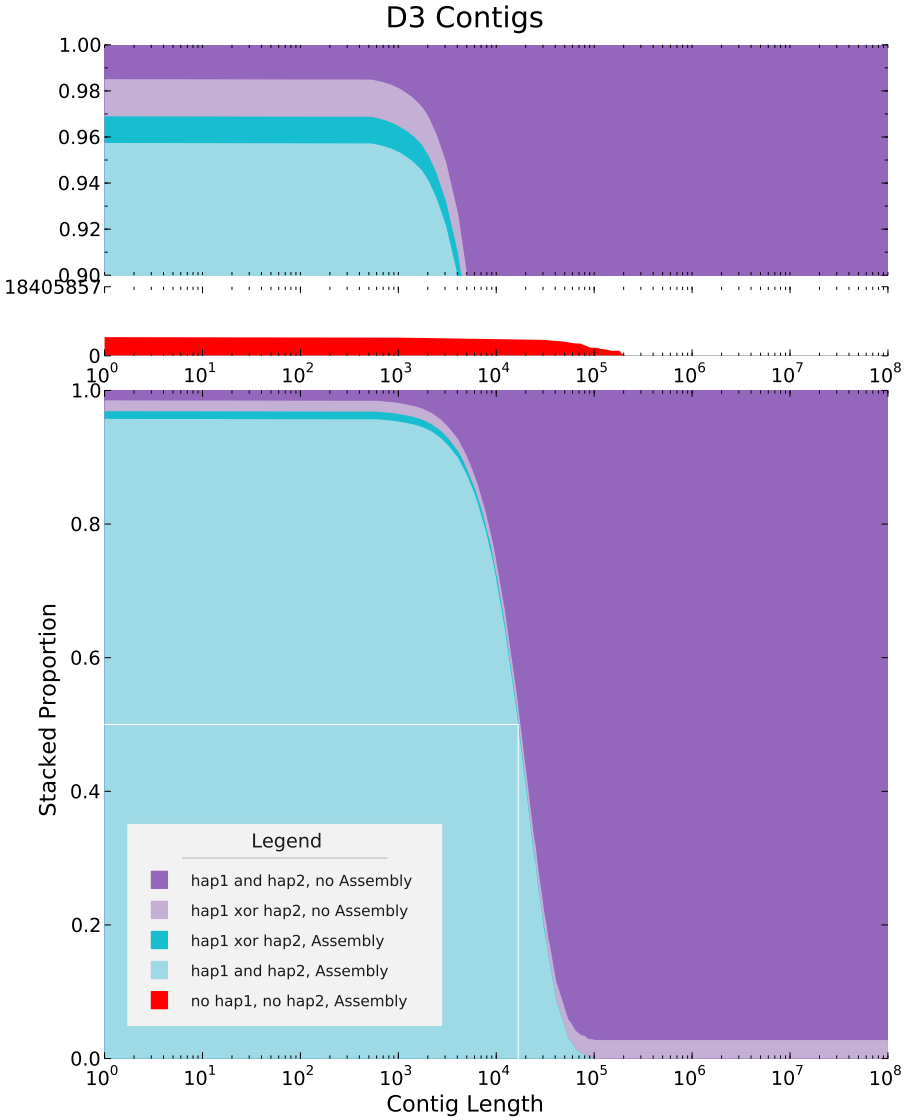


Figure 3.47: D3 contigs caption goes here.

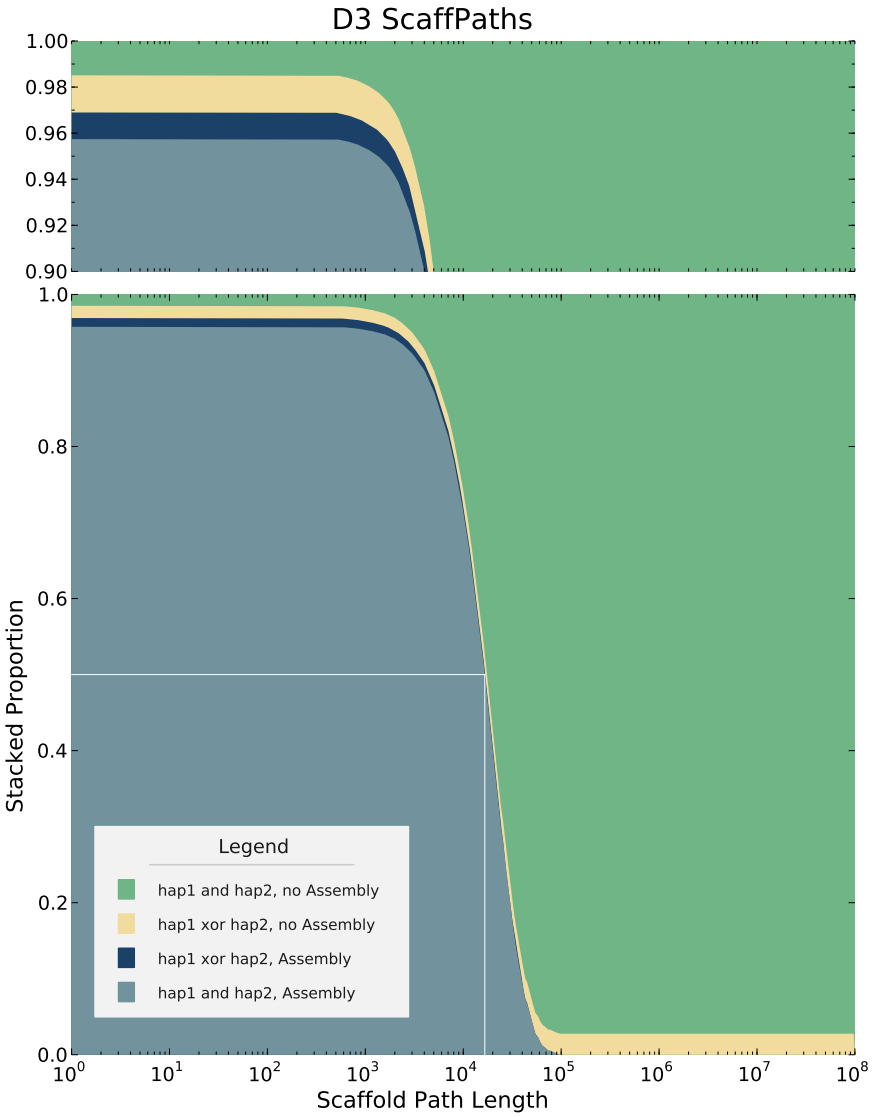


Figure 3.48: D3 scaffolds caption goes here.

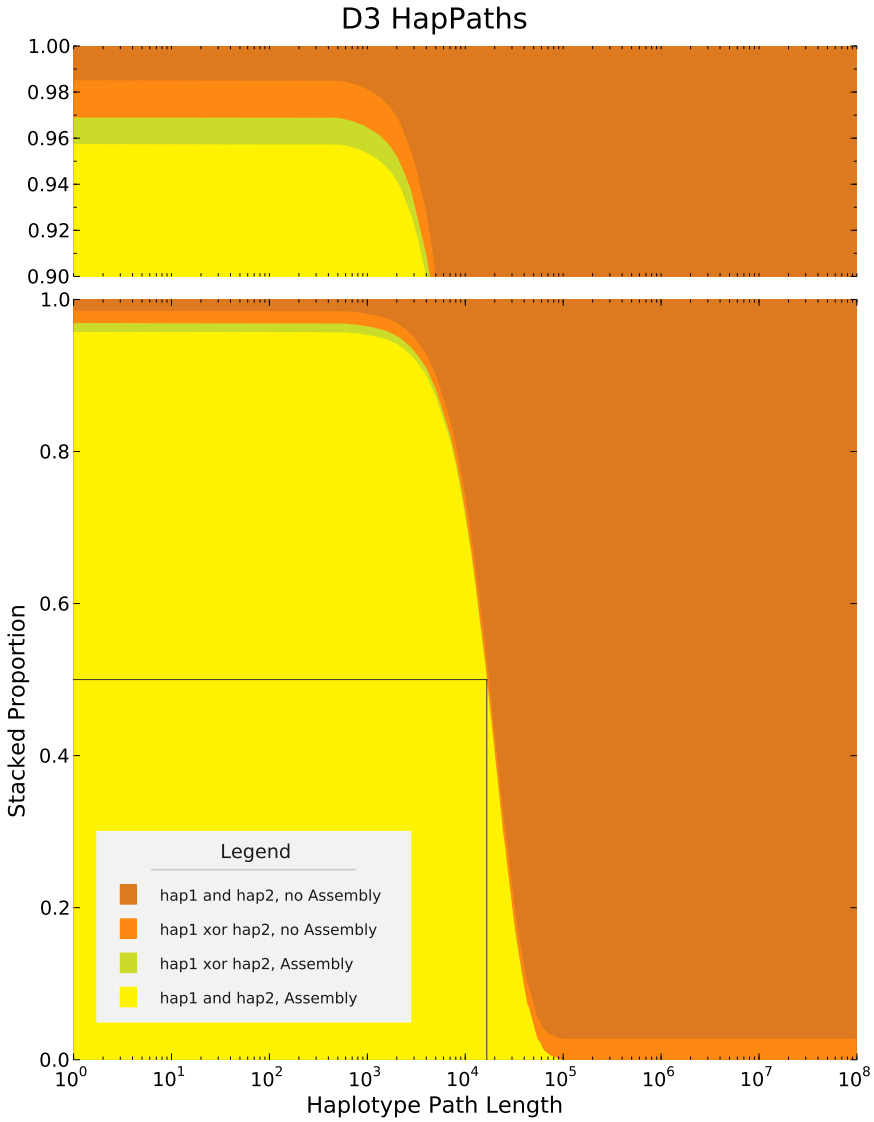


Figure 3.49: D3 hapPaths caption goes here.

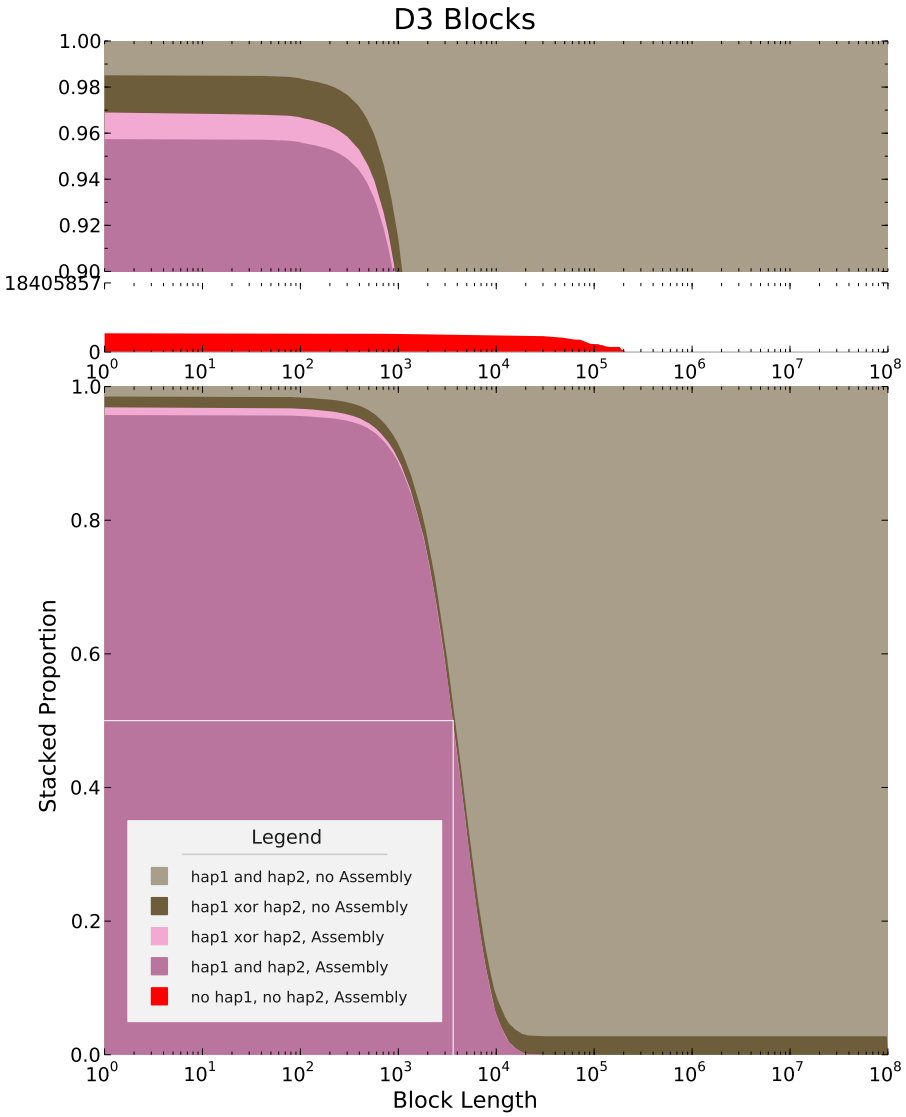


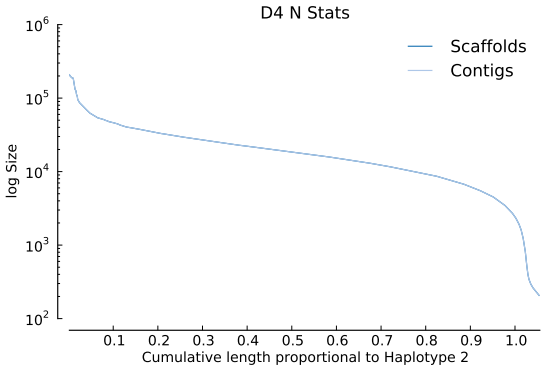
Figure 3.50: D3 blocks caption goes here.

D4

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
K1	0.98306	0.98309	0.98302	0.11258
D4	0.98288	0.98303	0.98274	0.99618
K2	0.98288	0.98287	0.98288	0.04786

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	23,037	200	246.00	386	5,156.38	6,443.00	207,957	9,761.96	118,787,544
Contigs	23,037	200	246.00	386	5,156.38	6,443.00	207,957	9,761.96	118,787,544

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	107,999,138 – 108,677,636	107,013,548 – 107,600,740	214,027,080.0 – 215,201,444.0	8 – 18
Heterozygous	421,001 – 432,419	416,034 – 423,813	832,068.0 – 847,626.0	0 – 0
Indel	2,791,469 – 3,187,403	1,275,033 – 1,548,458	2,546,814.0 – 3,093,534.0	1,626 – 1,691

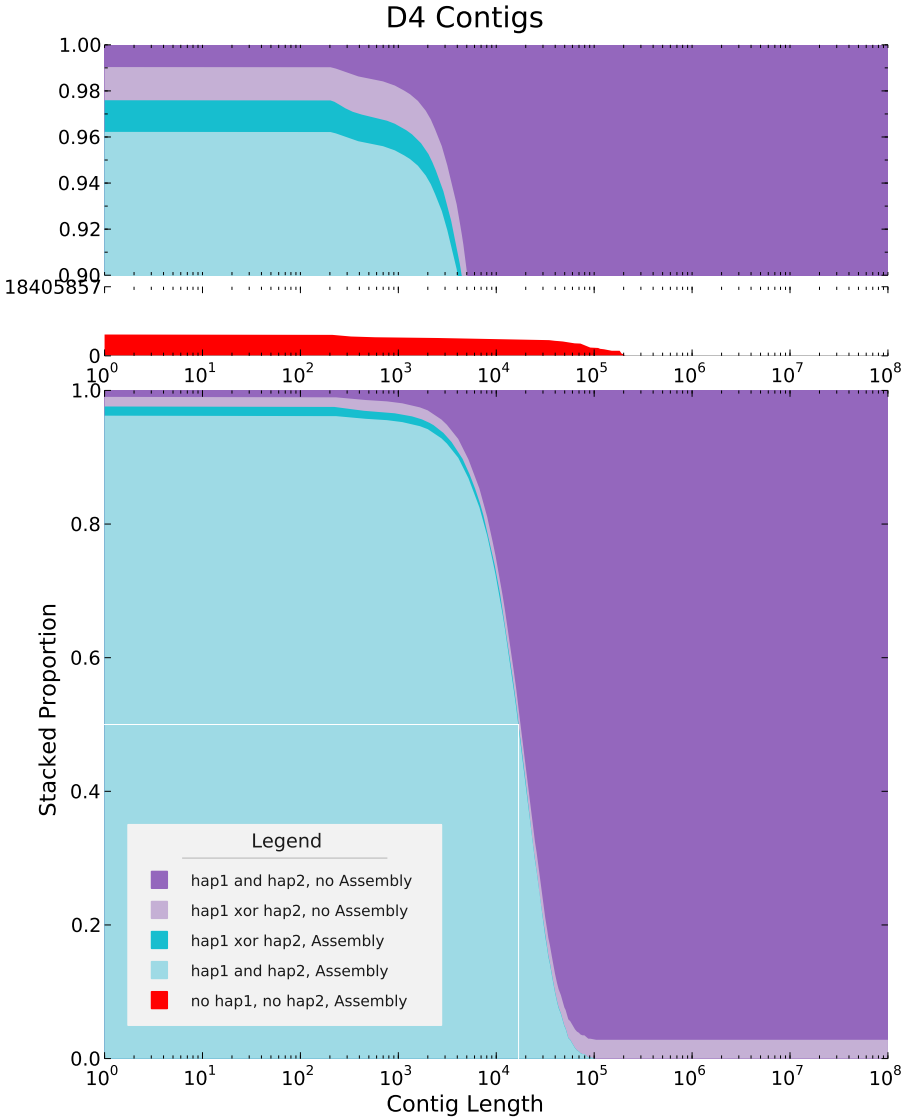


Figure 3.51: D4 contigs caption goes here.

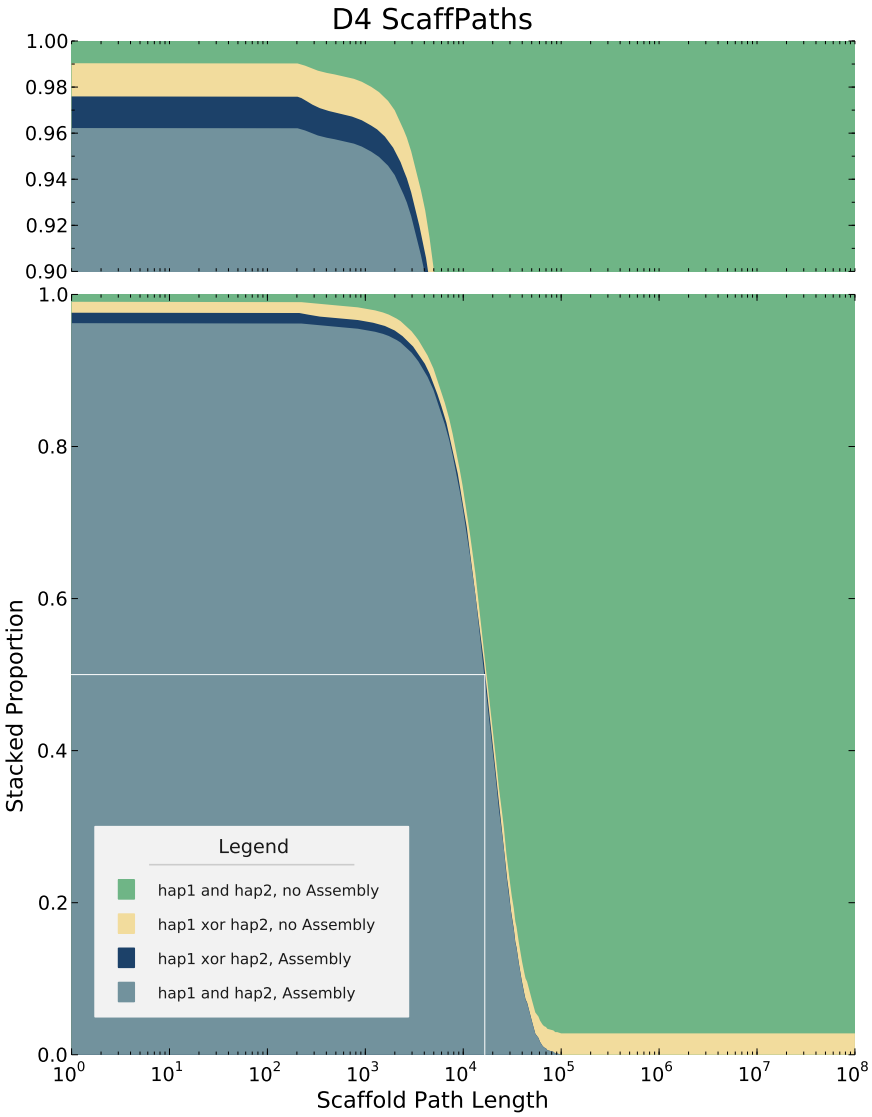


Figure 3.52: D4 scaffolds caption goes here.

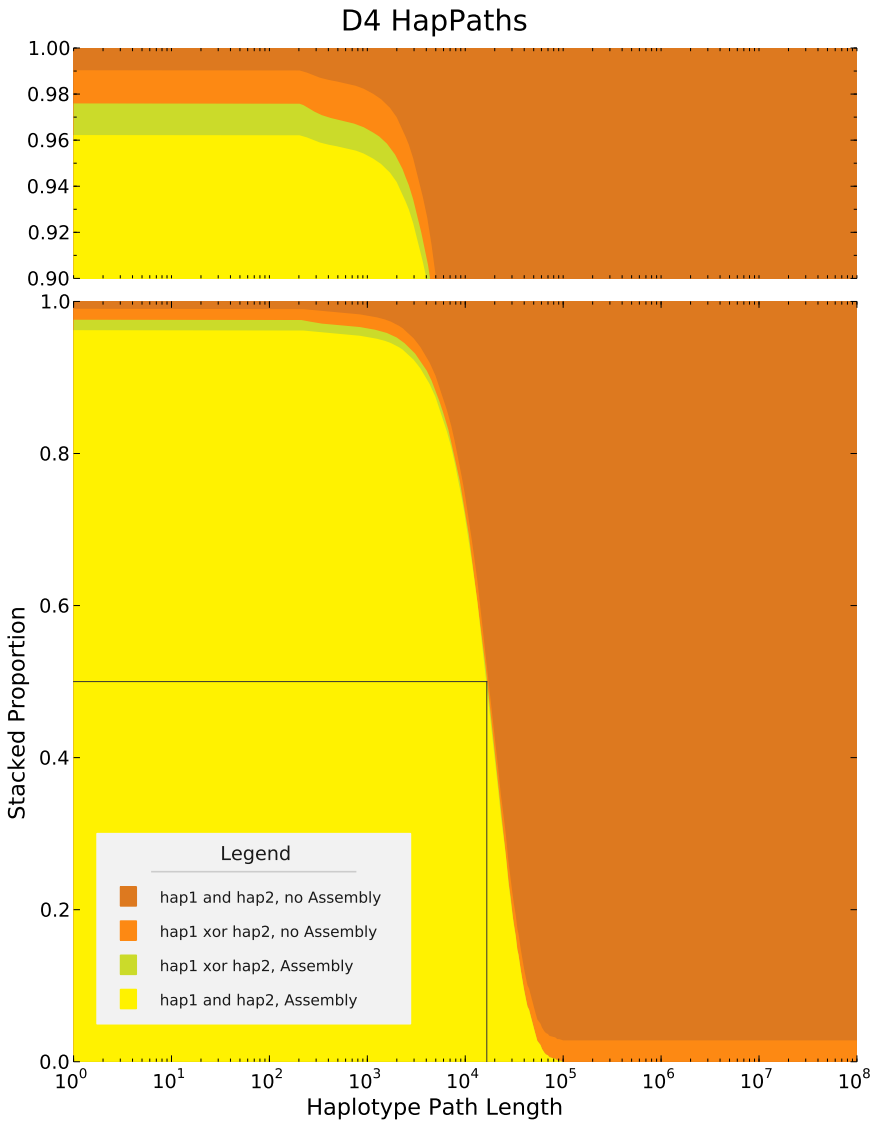


Figure 3.53: D4 hapPaths caption goes here.

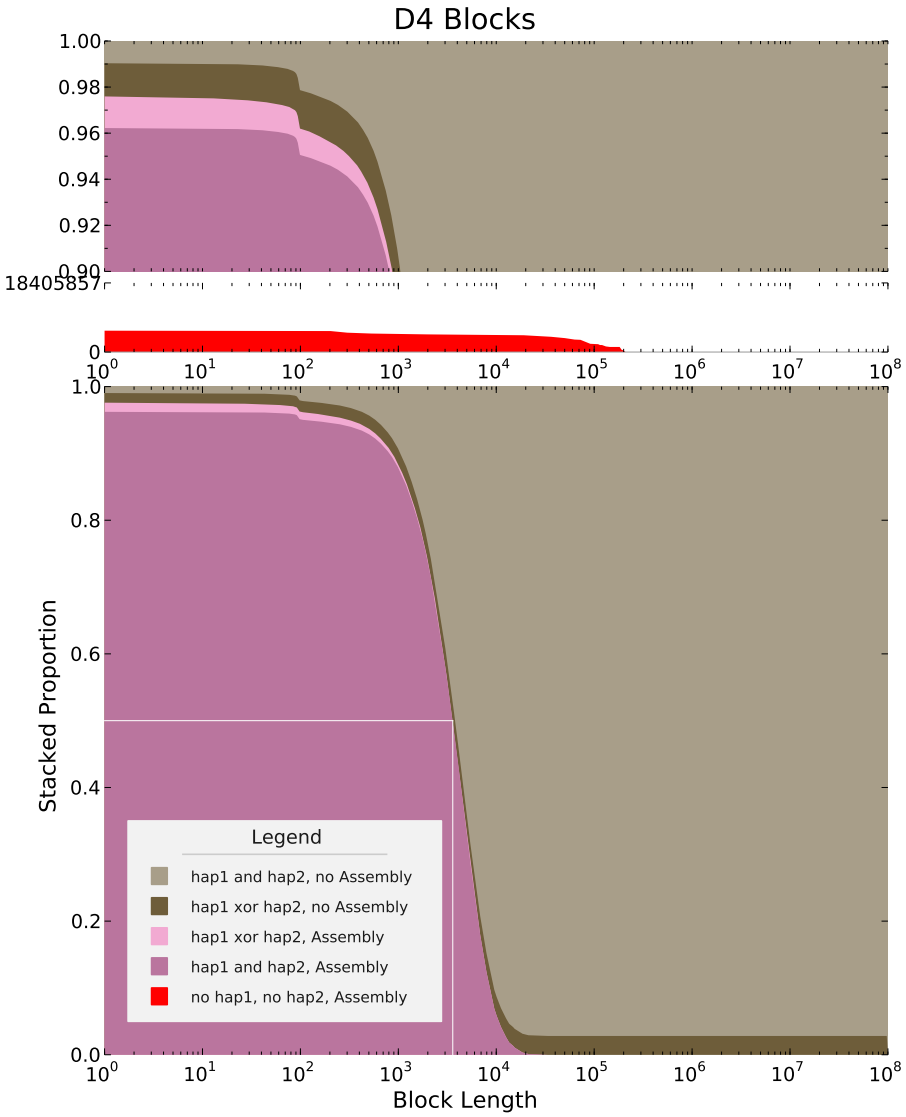


Figure 3.54: D4 blocks caption goes here.

3.2.5 E, Borgs

Affiliation: CRACS (Center for Research in Advanced Computing Systems), Portugal
Contact: Nuno Fonseca
Software: **ABySS**
Number of entries: 3

ID	Total	Hap 1	Hap 2	Bac
E1	0.96089	0.96125	0.96053	0.99725
E2	0.95596	0.95640	0.95551	0.99215
E3	0.95559	0.95601	0.95517	0.99317

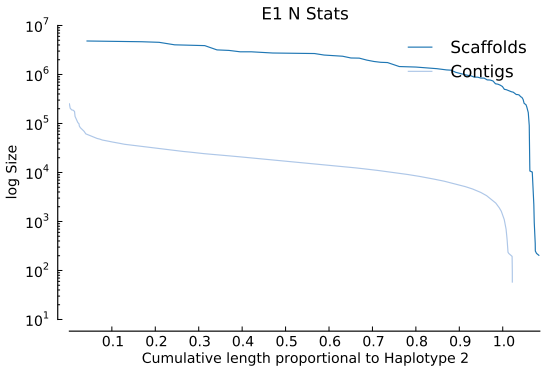
Assemblies:

E1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
V4	0.96153	0.96180	0.96128	0.99789
E1	0.96089	0.96125	0.96053	0.99725
E2	0.95596	0.95640	0.95551	0.99215

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	6,166	200	206.00	215	19,808.38	232.00	4,822,698	225,343.76	122,138,492
Contigs	17,341	57	221.00	2,510	6,632.91	9,449.00	251,891	10,438.23	115,021,309

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,234,018 – 109,183,195	105,036,662 – 105,666,068	210,071,100.0 – 211,328,142.0	42 – 100
Heterozygous	408,595 – 429,530	394,263 – 403,066	785,554.0 – 802,530.0	0 – 2
Indel	2,676,655 – 3,062,217	1,205,446 – 1,443,335	2,407,452.0 – 2,881,972.0	1,647 – 1,739

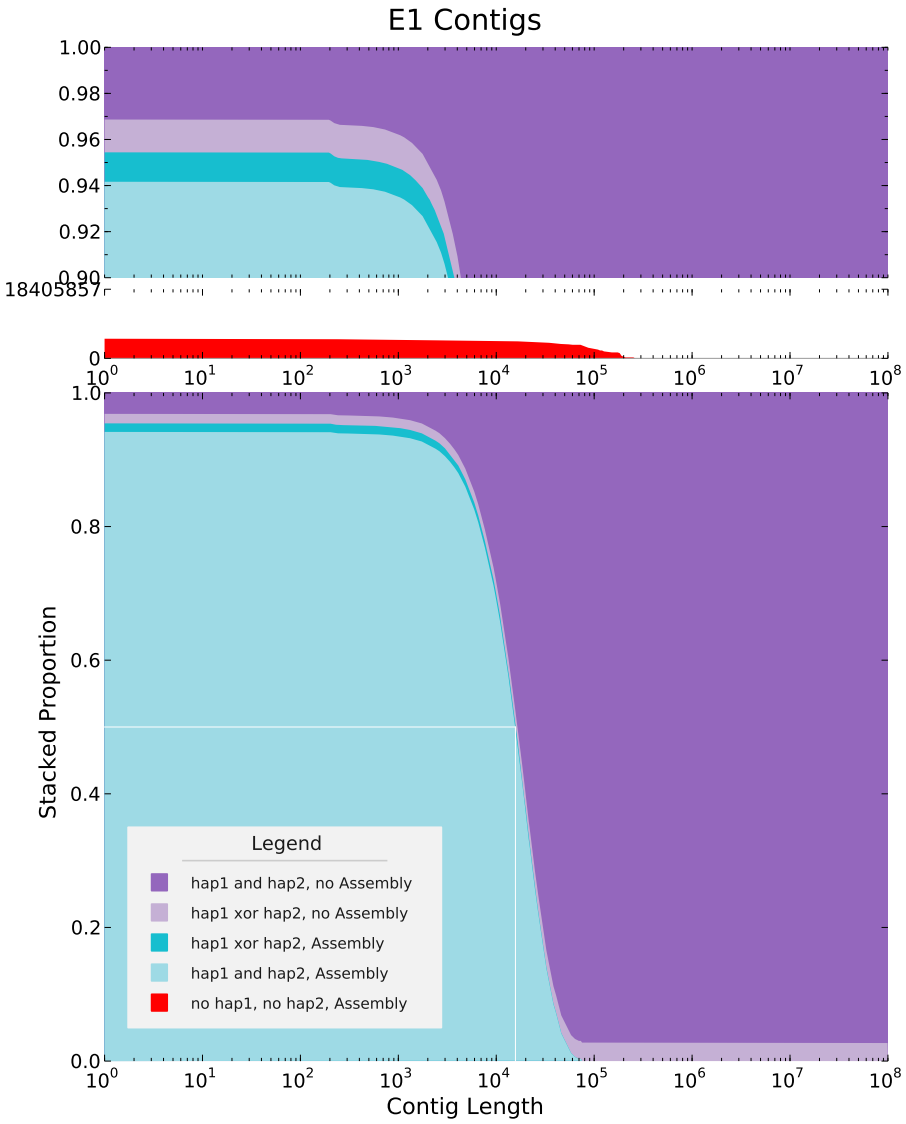


Figure 3.55: E1 contigs caption goes here.

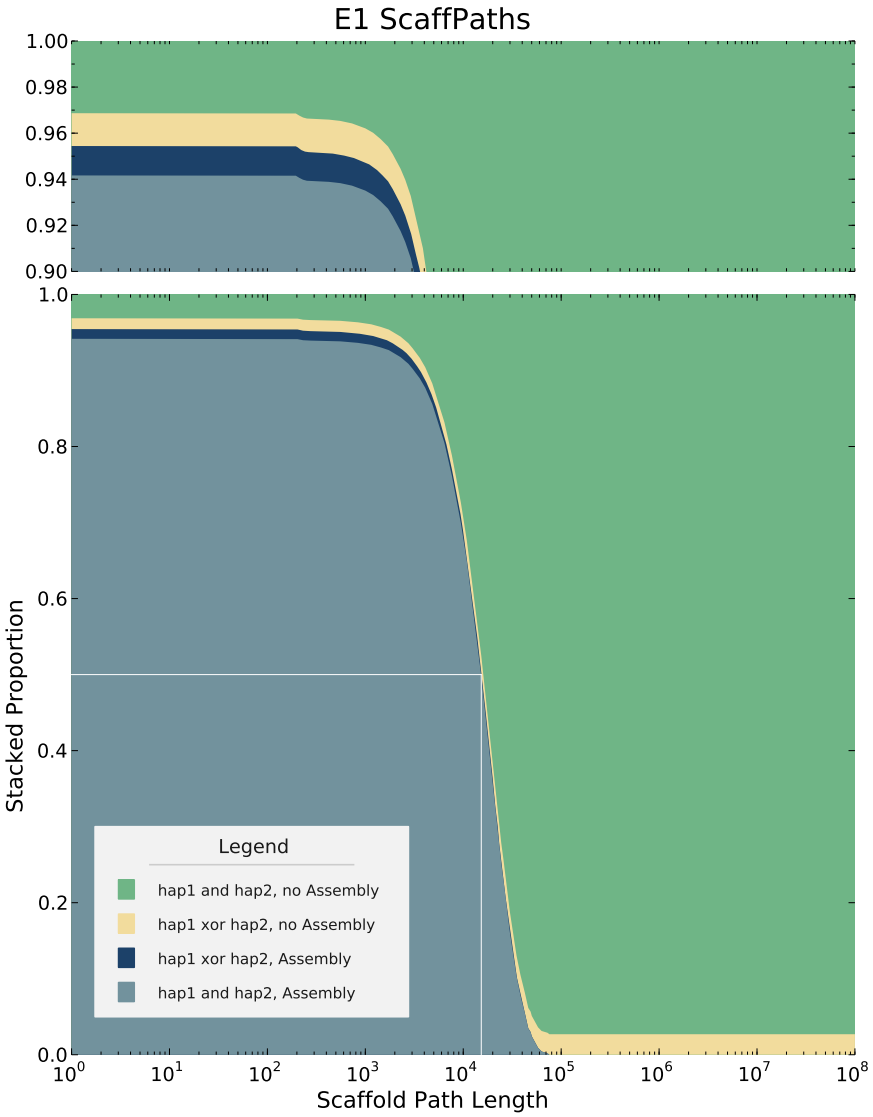


Figure 3.56: E1 scaffolds caption goes here.

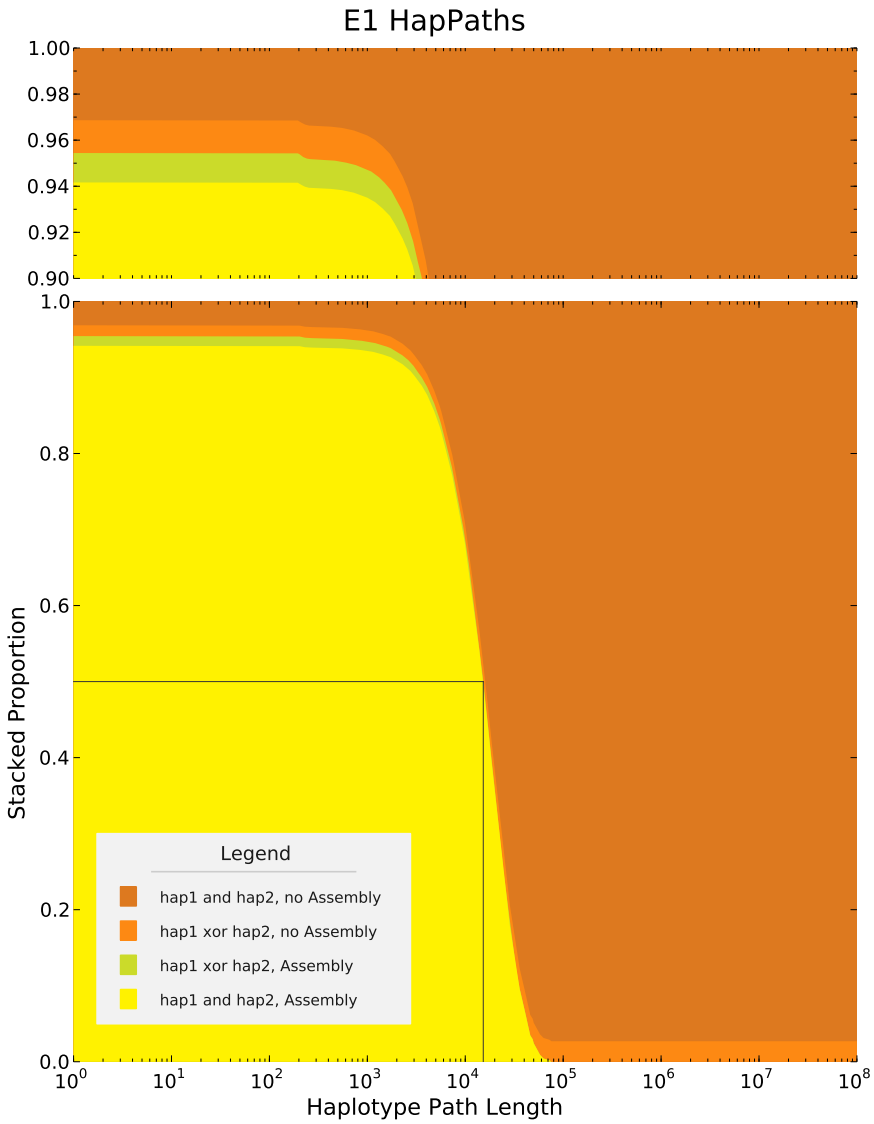


Figure 3.57: E1 hapPaths caption goes here.

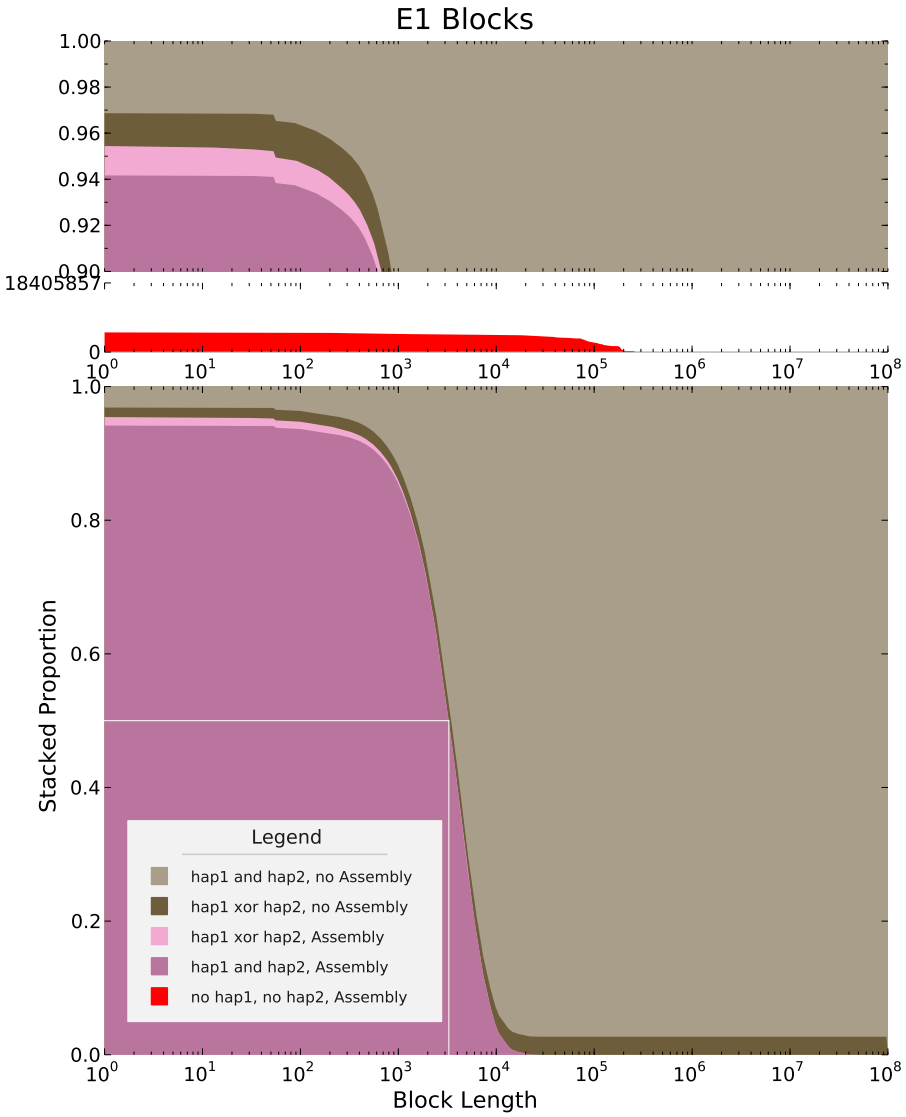


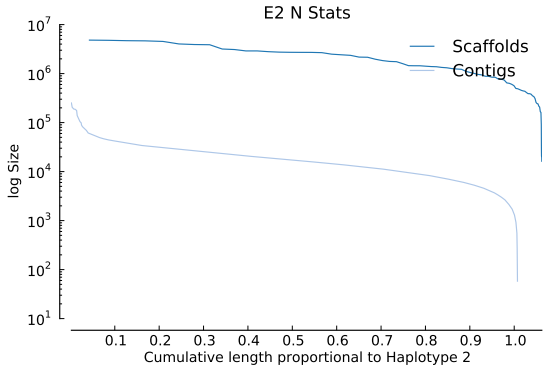
Figure 3.58: E1 blocks caption goes here.

E2

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
E1	0.96089	0.96125	0.96053	0.99725
E2	0.95596	0.95640	0.95551	0.99215
H4	0.95589	0.95589	0.95587	0.99681

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	82	16,258	416,121.00	1,032,300	1,457,686.71	2,311,142.75	4,822,698	1,320,439.92	119,530,310
Contigs	11,093	57	2,930.00	6,919	10,219.16	13,971.00	251,891	11,602.35	113,361,086

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,756,923 – 109,785,776	105,167,084 – 105,765,087	210,331,968.0 – 211,526,330.0	41 – 100
Heterozygous	413,233 – 438,639	395,795 – 404,142	788,636.0 – 804,734.0	0 – 2
Indel	2,662,361 – 3,044,670	1,191,904 – 1,404,952	2,380,370.0 – 2,805,176.0	1,643 – 1,733

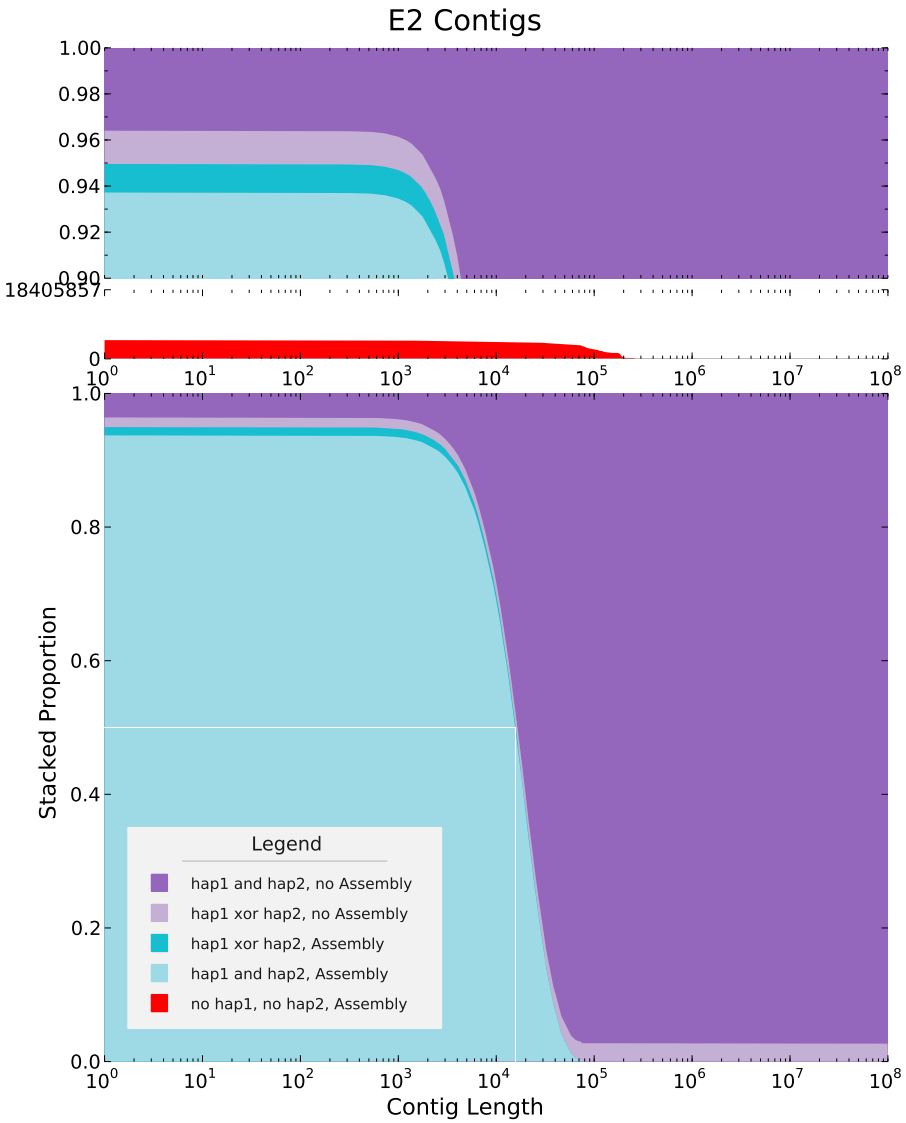


Figure 3.59: E2 contigs caption goes here.

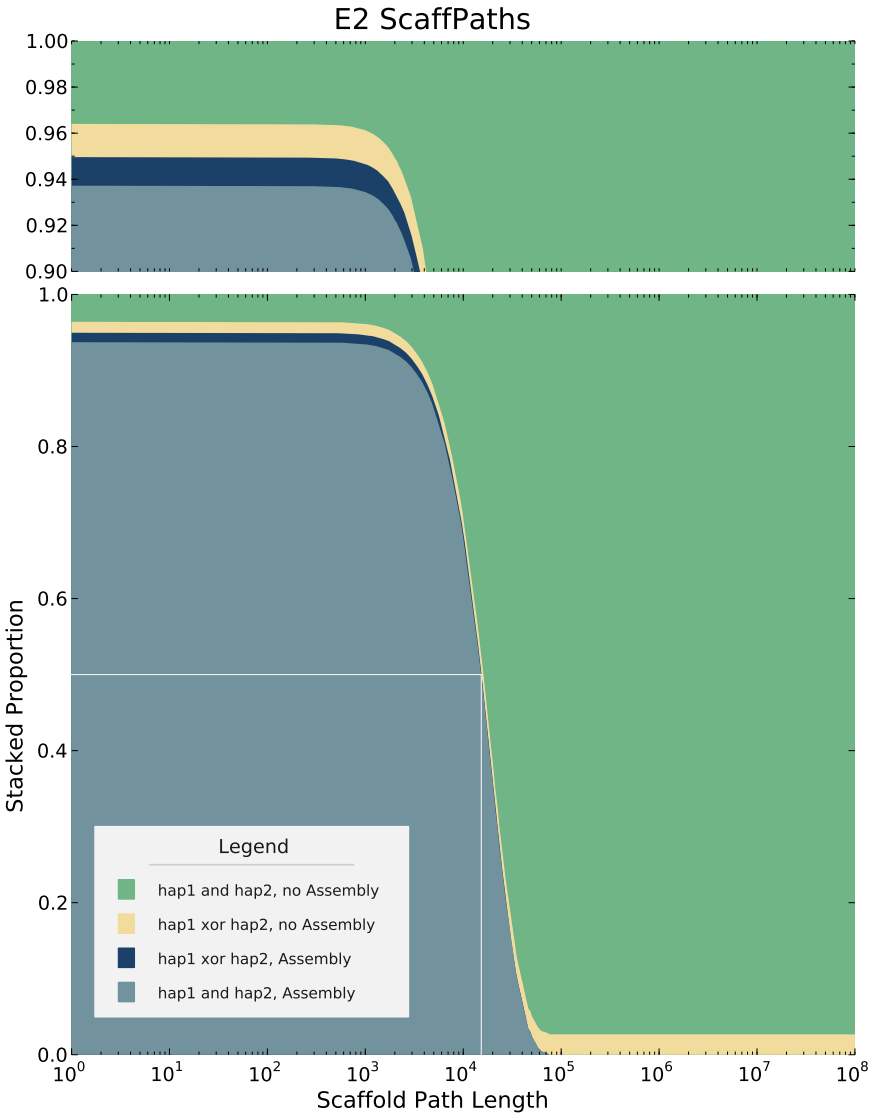


Figure 3.60: E2 scaffolds caption goes here.

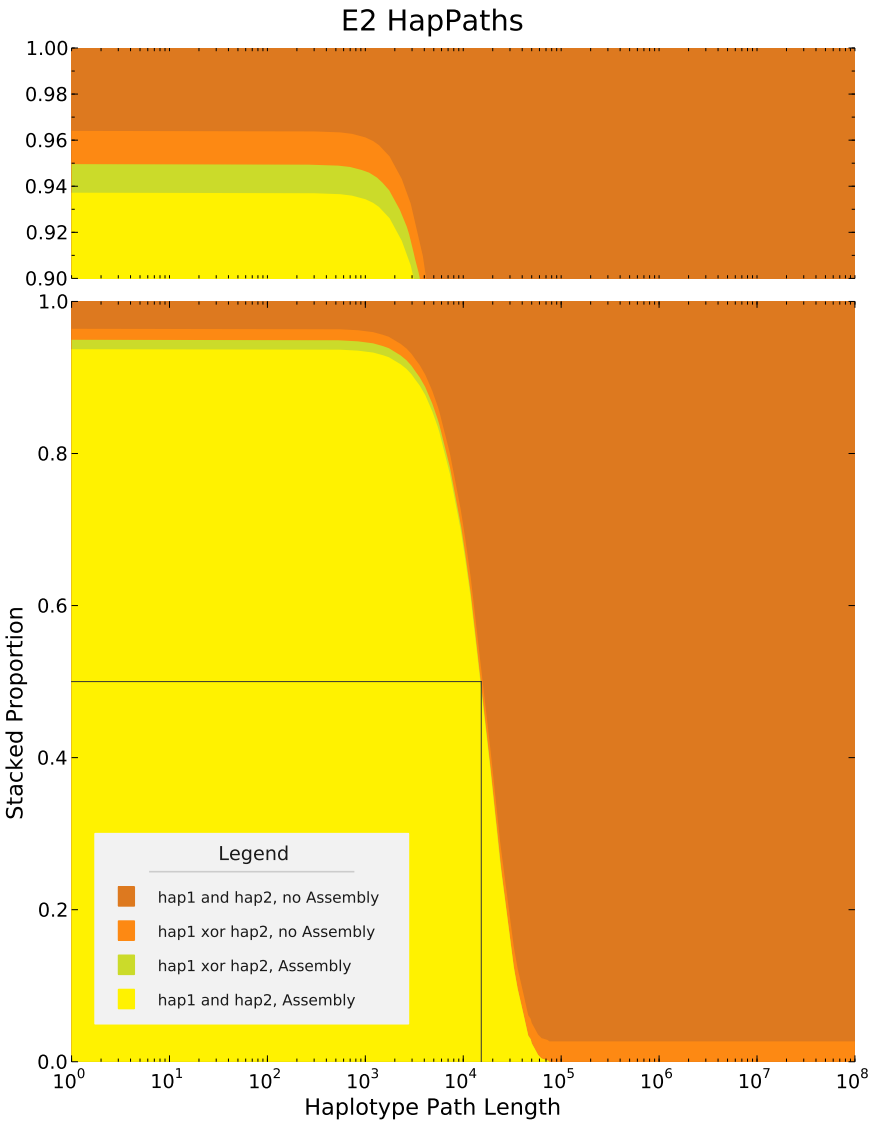


Figure 3.61: E2 hapPaths caption goes here.

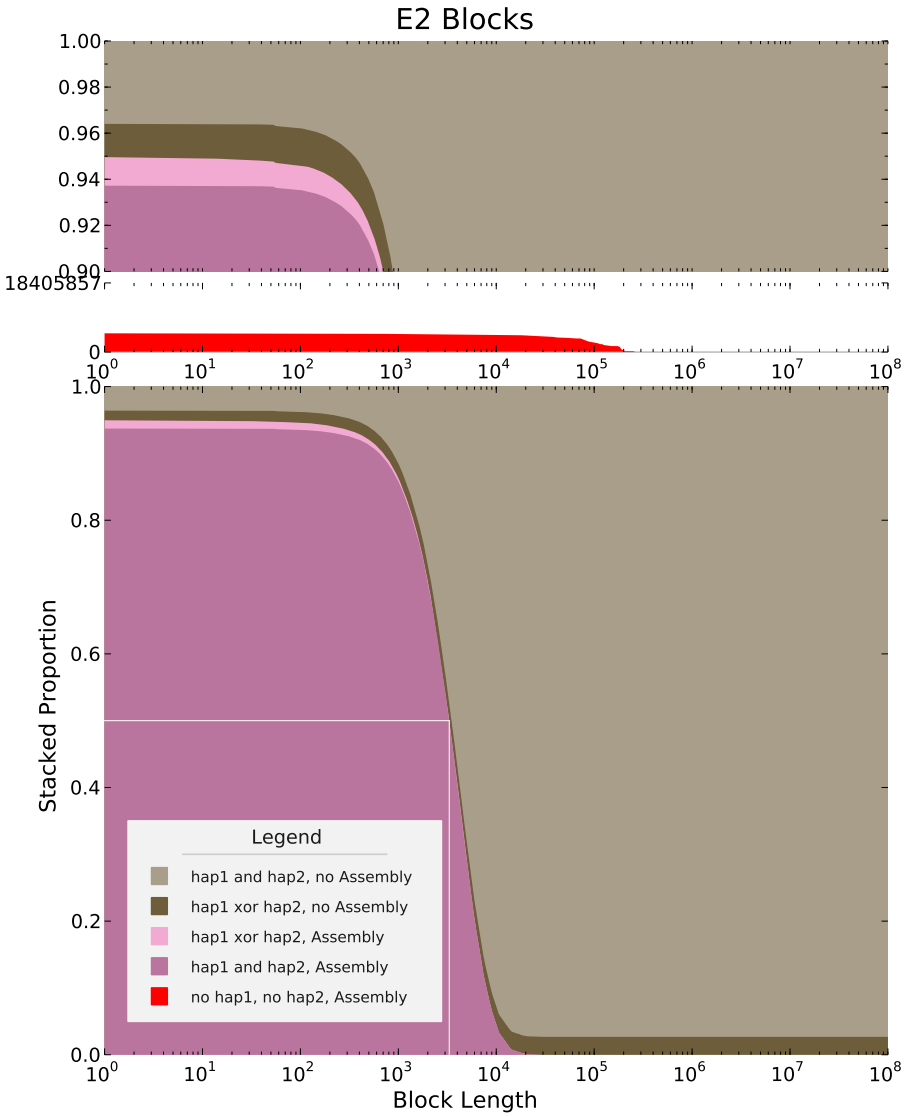


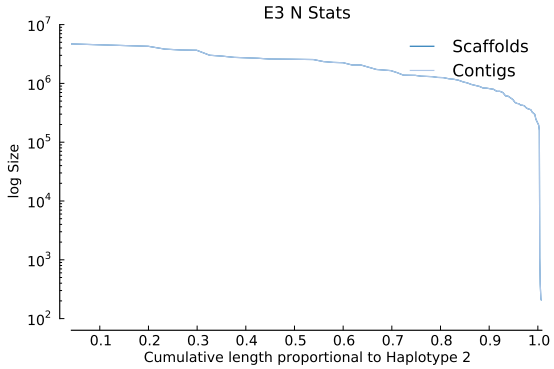
Figure 3.62: E2 blocks caption goes here.

E3

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
H4	0.95589	0.95589	0.95587	0.99681
E3	0.95559	0.95601	0.95517	0.99317
X6	0.95516	0.95527	0.95504	0.99562

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	1,937	200	207.00	217	58,568.72	236.00	4,689,666	379,195.82	113,447,607
Contigs	1,937	200	207.00	217	58,568.72	236.00	4,689,666	379,195.82	113,447,607

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,775,955 – 109,854,286	104,968,462 – 105,708,062	209,936,368.0 – 211,410,236.0	278 – 2,944
Heterozygous	412,621 – 439,130	395,697 – 409,394	791,358.0 – 817,566.0	18 – 611
Indel	2,488,705 – 2,873,311	1,146,992 – 1,363,595	2,290,938.0 – 2,722,954.0	1,523 – 2,118

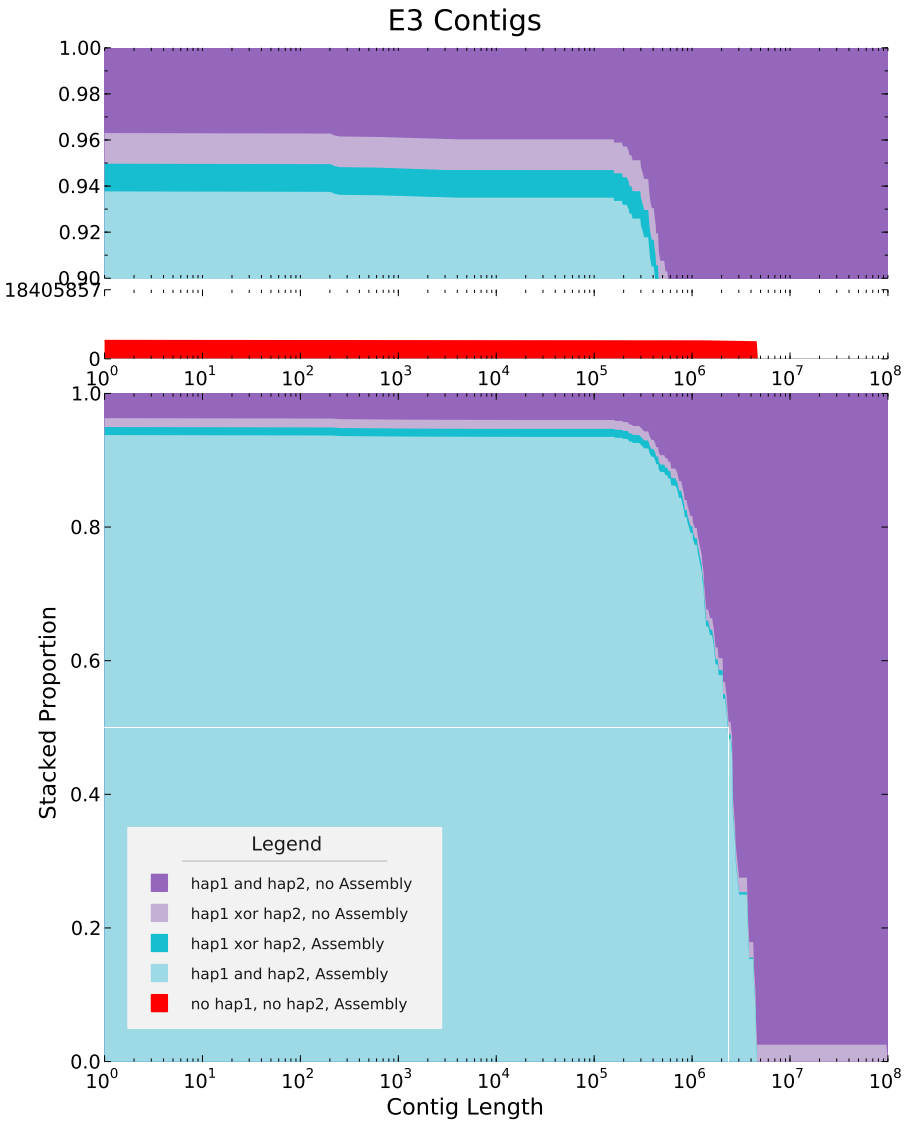


Figure 3.63: E3 contigs caption goes here.

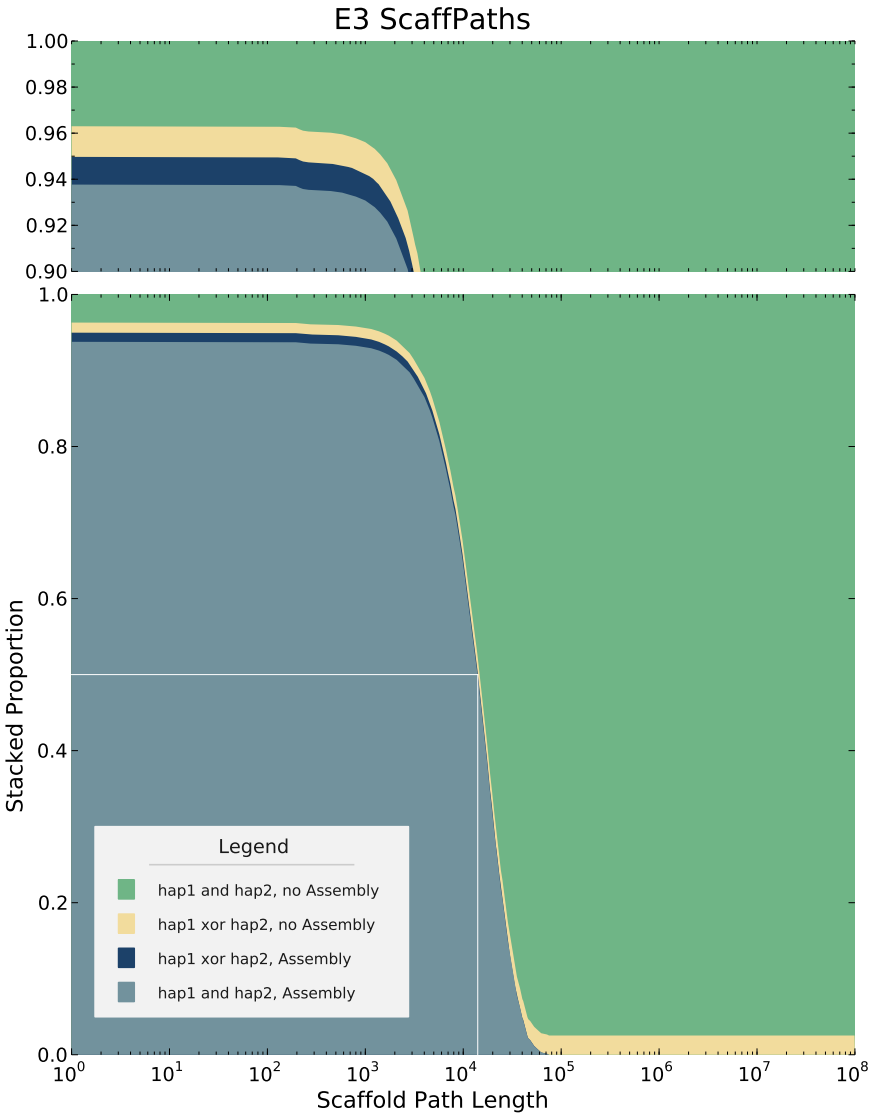


Figure 3.64: E3 scaffolds caption goes here.

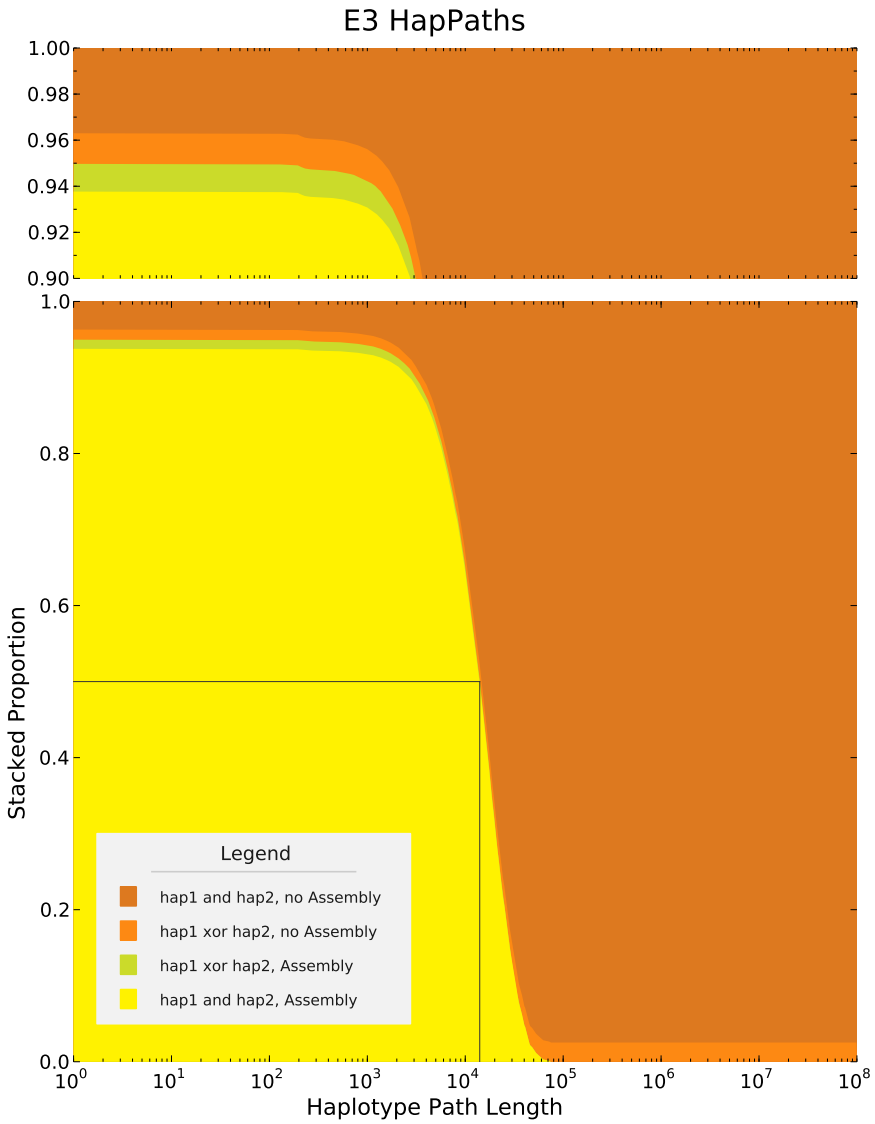


Figure 3.65: E3 hapPaths caption goes here.

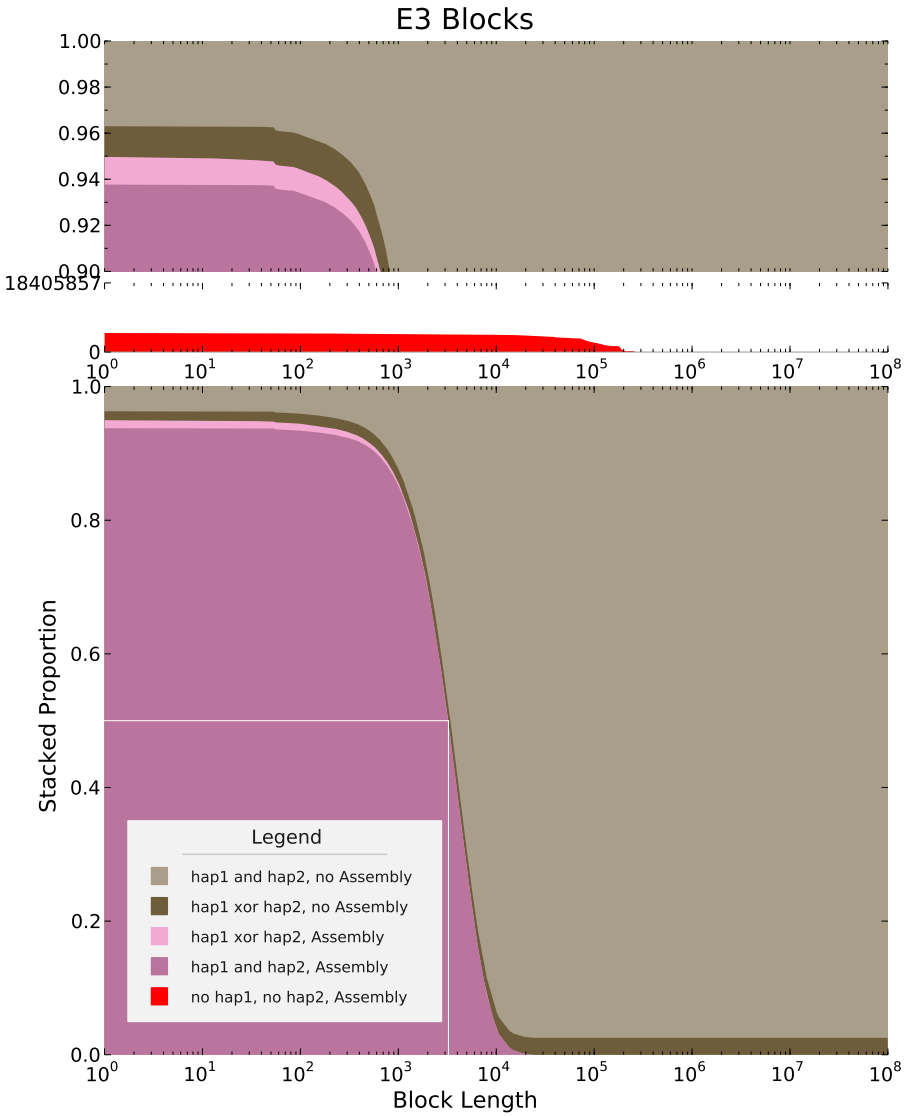


Figure 3.66: E3 blocks caption goes here.

3.2.6 F, ABySS

Affiliation: BC Cancer Genome Sciences Centre, Canada
Contact: Shaun Jackman
Software: **ABySS, Anchor**
Number of entries: 5

ID	Total	Hap 1	Hap 2	Bac
F5	0.98691	0.98727	0.98653	0.99934
F3	0.98671	0.98696	0.98648	0.99927
F4	0.98649	0.98675	0.98624	0.99928
F2	0.98644	0.98676	0.98611	0.99924
F1	0.98630	0.98664	0.98595	0.99923

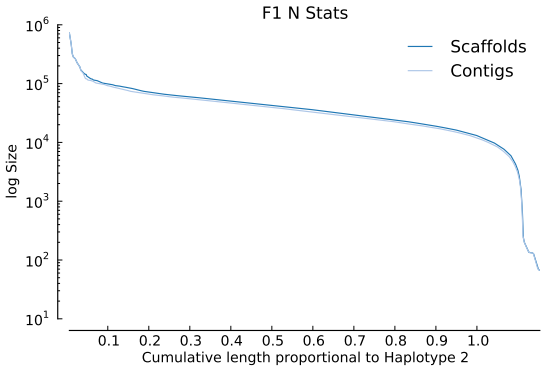
Assemblies:

F1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
F2	0.98644	0.98676	0.98611	0.99924
F1	0.98630	0.98664	0.98595	0.99923
B2	0.98568	0.98600	0.98535	0.99892

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	44,949	67	80.00	131	2,887.46	160.00	721,724	11,914.76	129,788,571
Contigs	45,487	67	80.00	131	2,852.80	163.00	721,724	11,436.56	129,765,436

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	99,990,258 – 100,565,939	99,304,668 – 99,834,763	198,608,578.4 – 199,668,465.4	73 – 111
Heterozygous	389,205 – 397,277	385,682 – 392,637	770,084.4 – 783,488.4	601 – 834
Indel	2,793,458 – 3,193,355	1,280,590 – 1,588,559	2,557,560.0 – 3,172,521.0	1,767 – 1,857

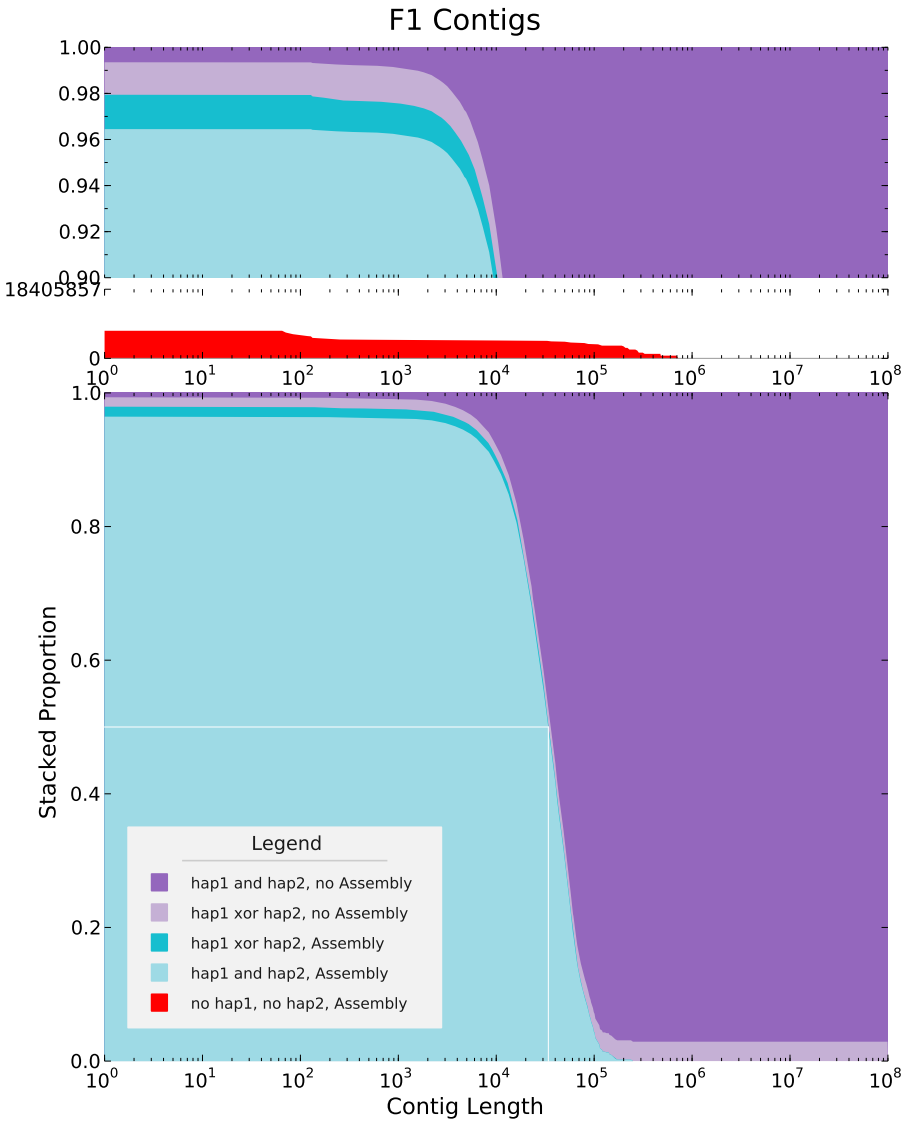


Figure 3.67: F1 contigs caption goes here.

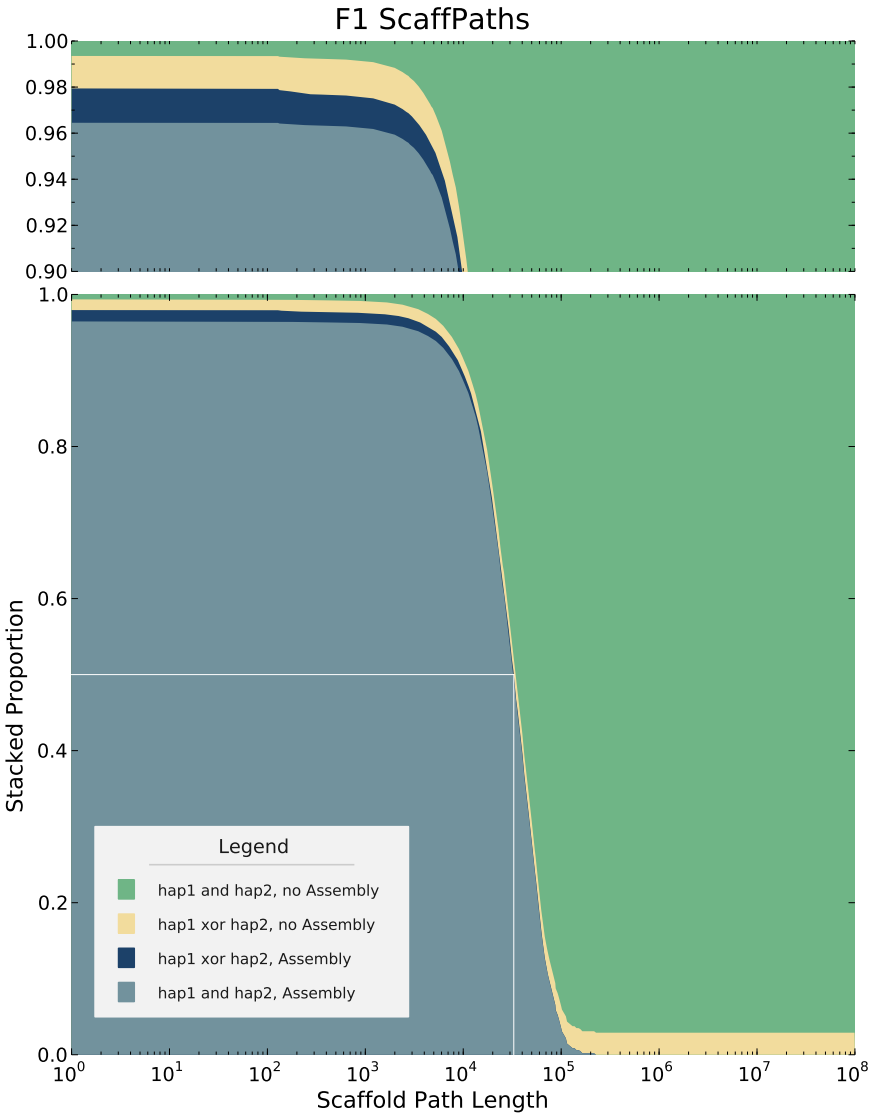


Figure 3.68: F1 scaffolds caption goes here.

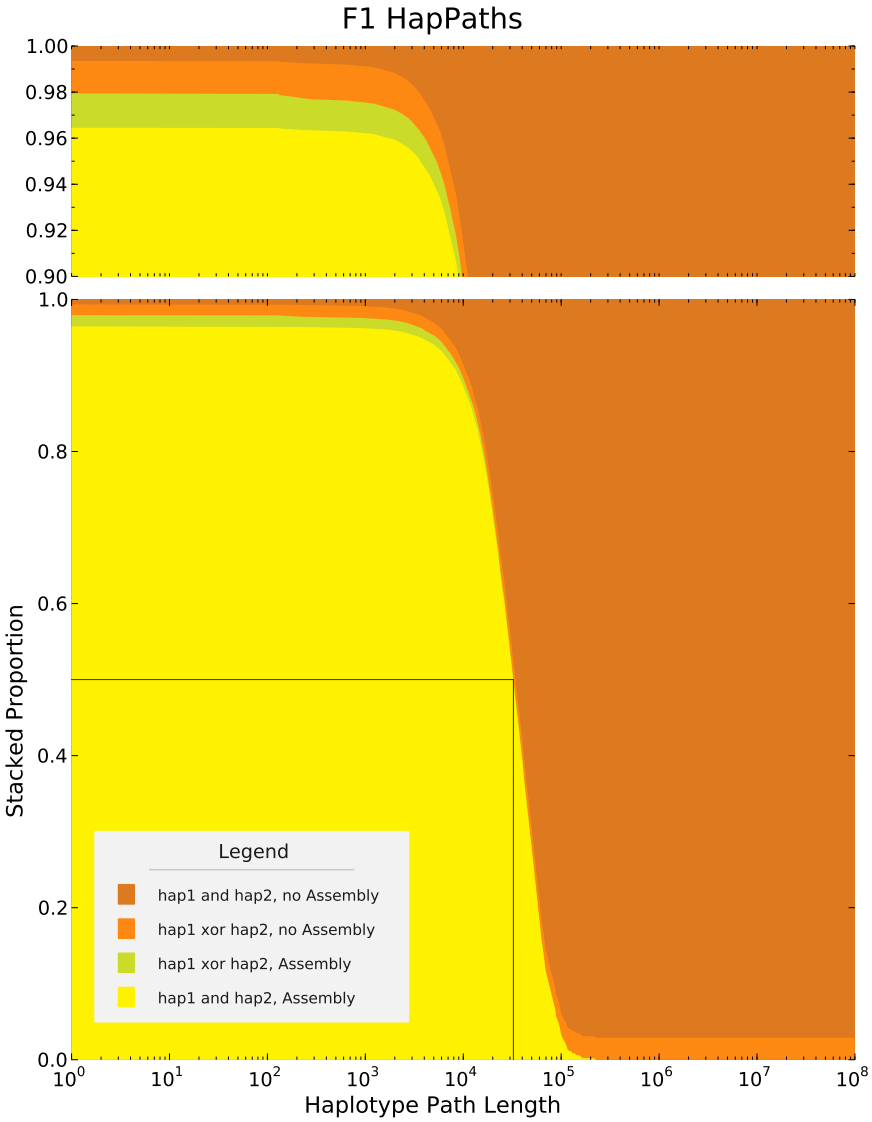


Figure 3.69: F1 hapPaths caption goes here.

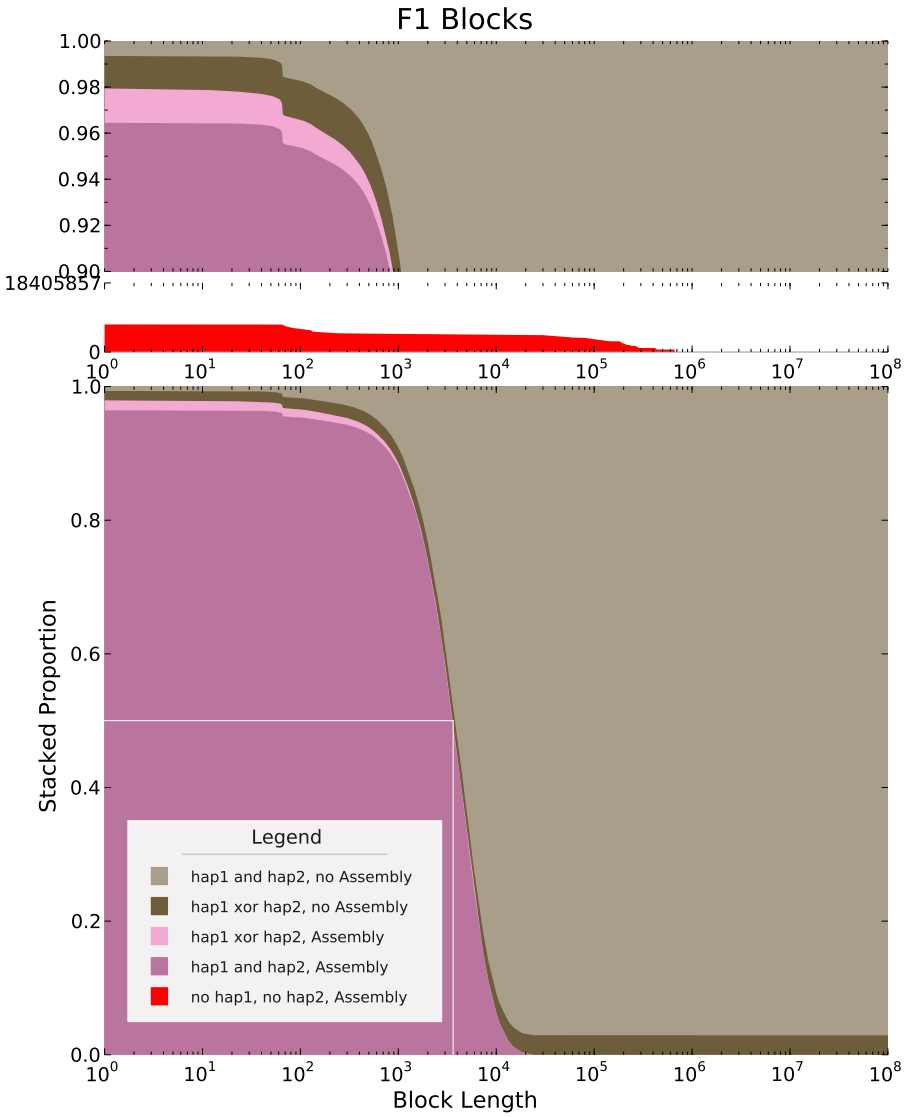


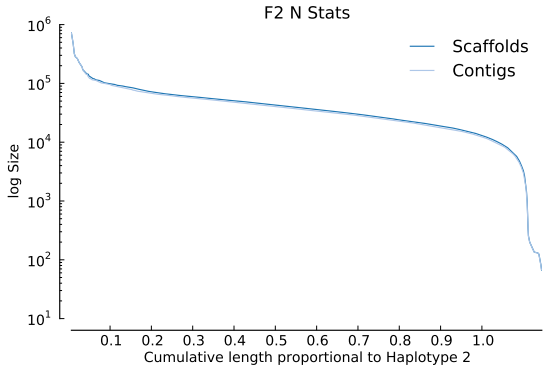
Figure 3.70: F1 blocks caption goes here.

F2

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
F4	0.98649	0.98675	0.98624	0.99928
F2	0.98644	0.98676	0.98611	0.99924
F1	0.98630	0.98664	0.98595	0.99923

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	33,063	67	122.50	133	3,897.06	189.00	721,724	13,752.93	128,848,474
Contigs	33,437	67	123.00	133	3,852.99	192.00	721,724	13,358.29	128,832,348

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	100,034,692 – 100,609,131	99,348,767 – 99,878,337	198,696,942.0 – 199,755,641.0	44 – 75
Heterozygous	389,653 – 397,558	386,121 – 392,964	769,830.0 – 783,076.0	1,186 – 1,390
Indel	2,758,123 – 3,162,489	1,272,405 – 1,584,568	2,541,202.0 – 3,164,760.0	1,784 – 1,913

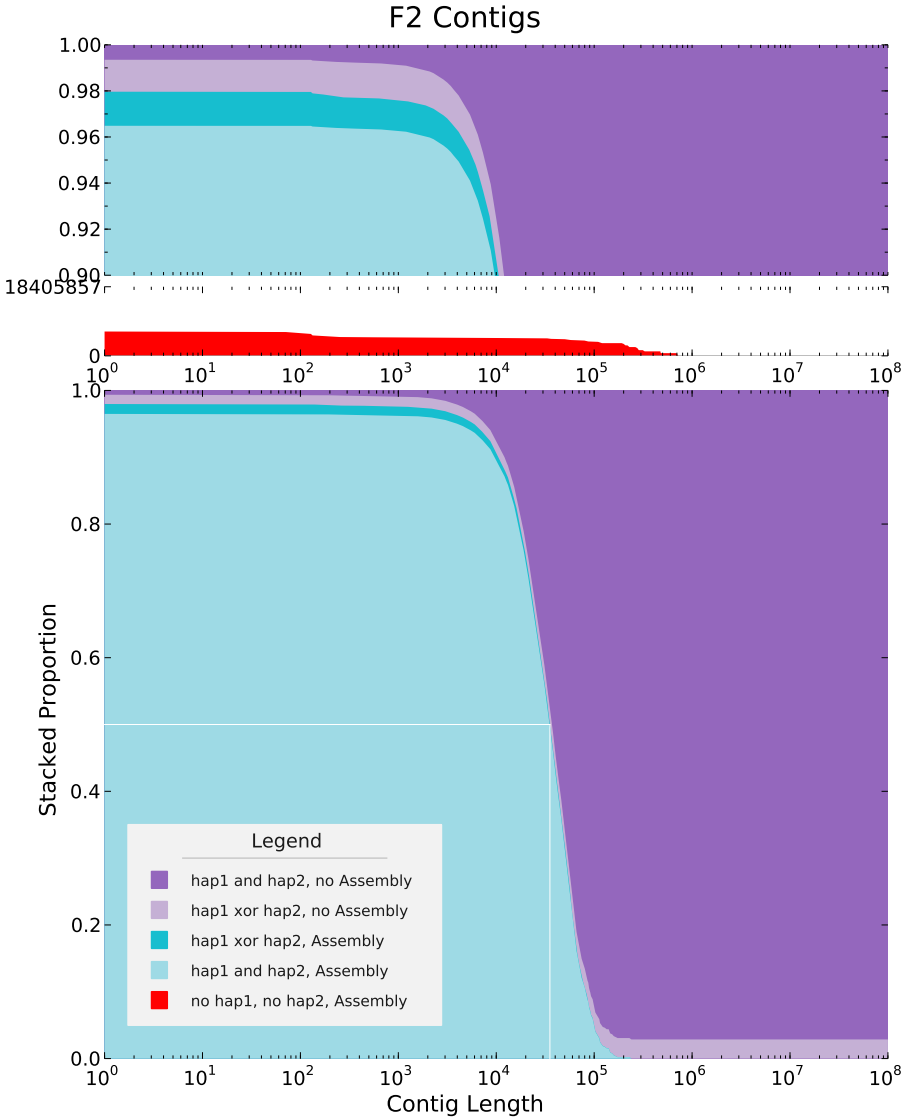


Figure 3.71: F2 contigs caption goes here.

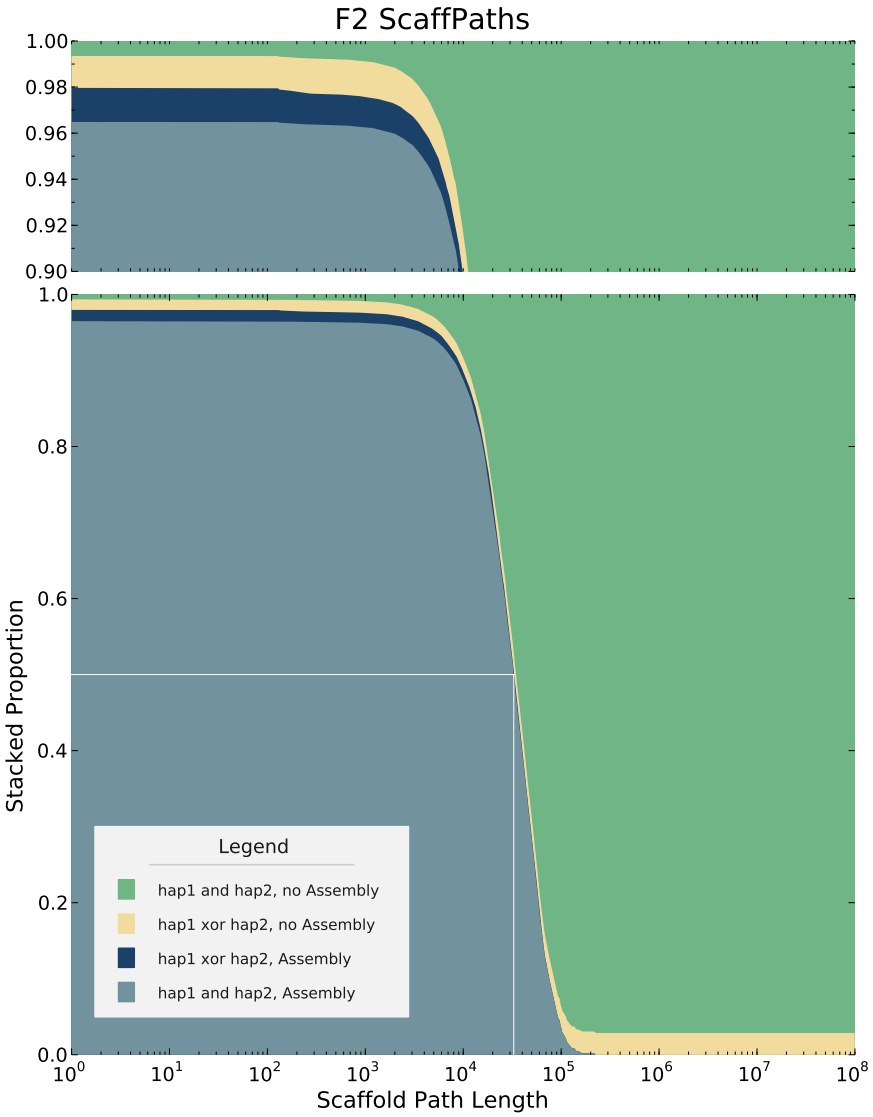


Figure 3.72: F2 scaffolds caption goes here.

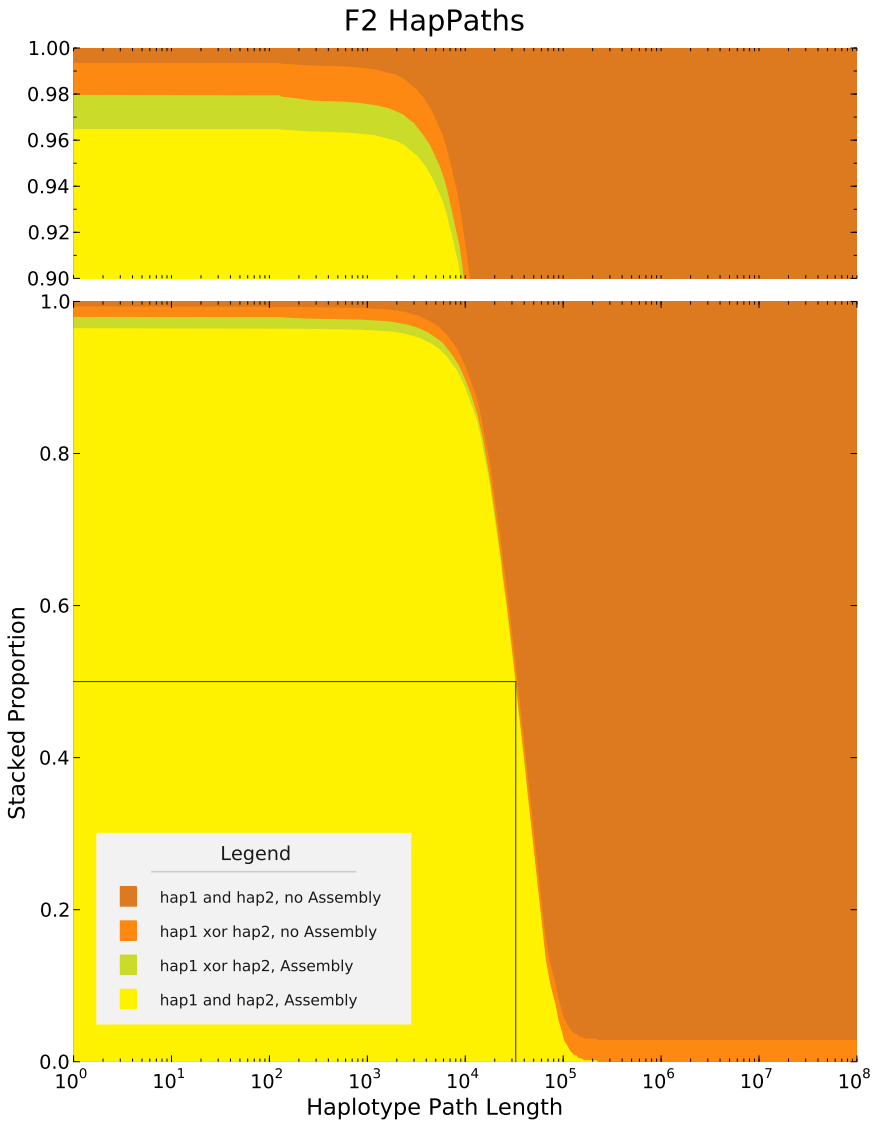


Figure 3.73: F2 hapPaths caption goes here.

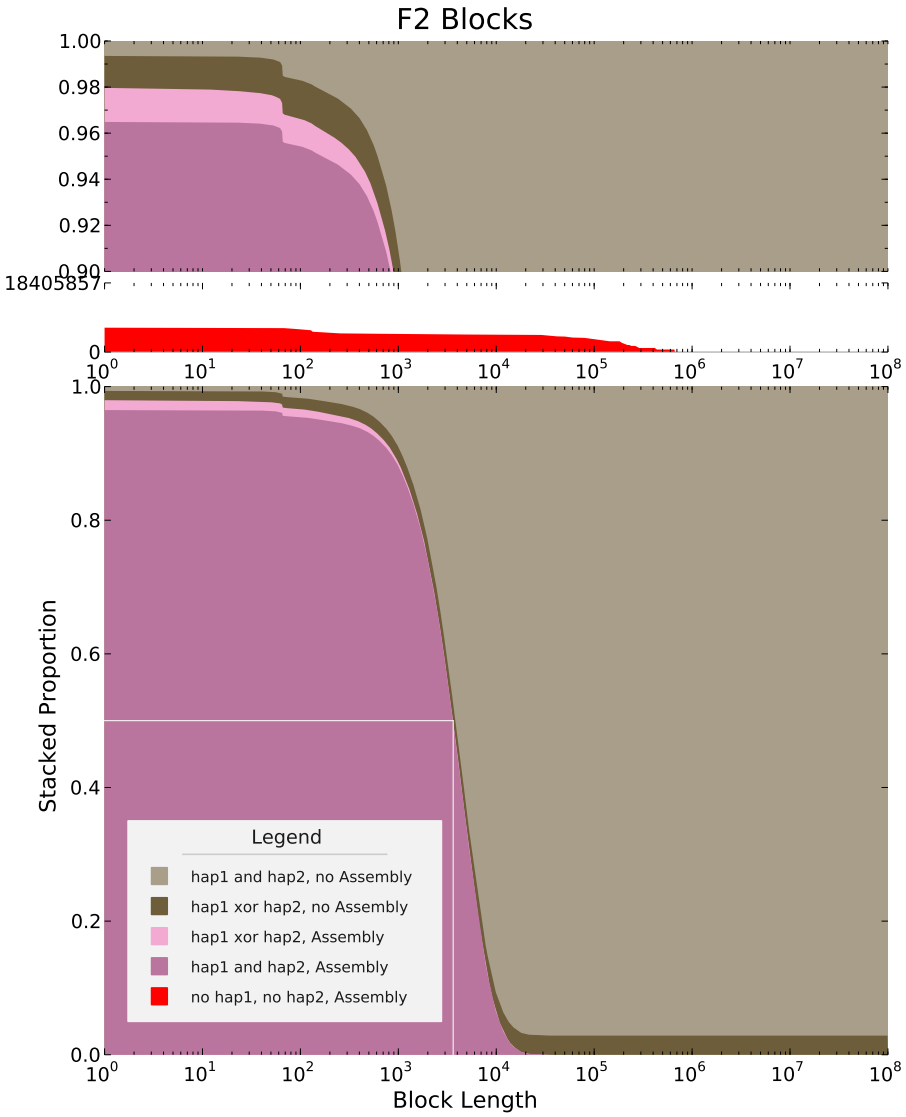


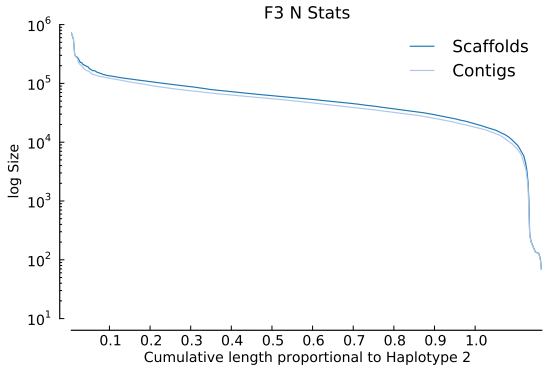
Figure 3.74: F2 blocks caption goes here.

F3

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
M3	0.98674	0.98685	0.98662	0.99977
F3	0.98671	0.98696	0.98648	0.99927
M5	0.98667	0.98679	0.98655	0.99969

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	28,678	67	117.00	133	4,567.99	183.00	721,864	17,416.86	131,000,911
Contigs	29,300	67	118.00	133	4,469.77	189.00	721,864	16,205.48	130,964,214

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	97,803,309 – 98,352,965	97,156,865 – 97,664,777	194,312,828.0 – 195,328,173.0	82 – 148
Heterozygous	381,474 – 389,288	378,136 – 385,060	752,884.0 – 766,166.0	1,662 – 1,906
Indel	2,779,611 – 3,178,046	1,276,989 – 1,583,179	2,550,377.0 – 3,161,593.0	1,726 – 1,901

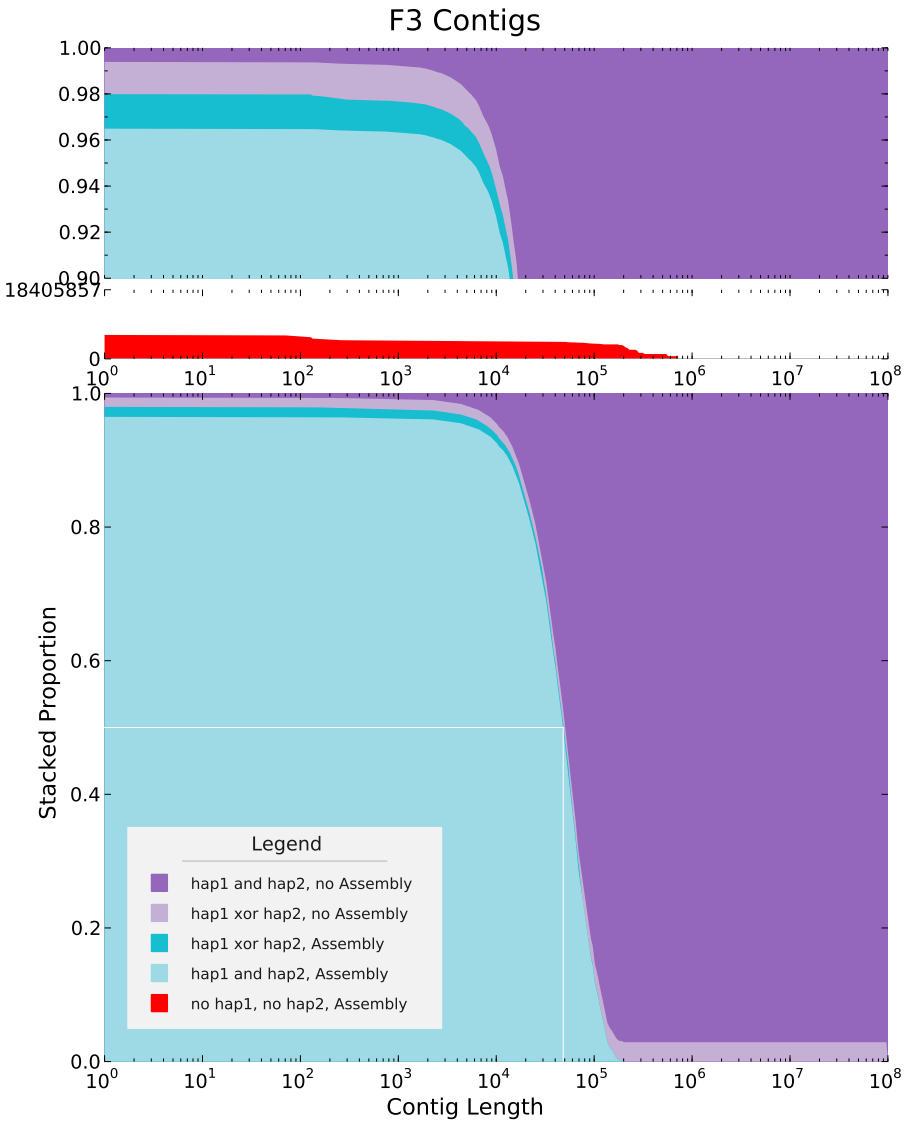


Figure 3.75: F3 contigs caption goes here.

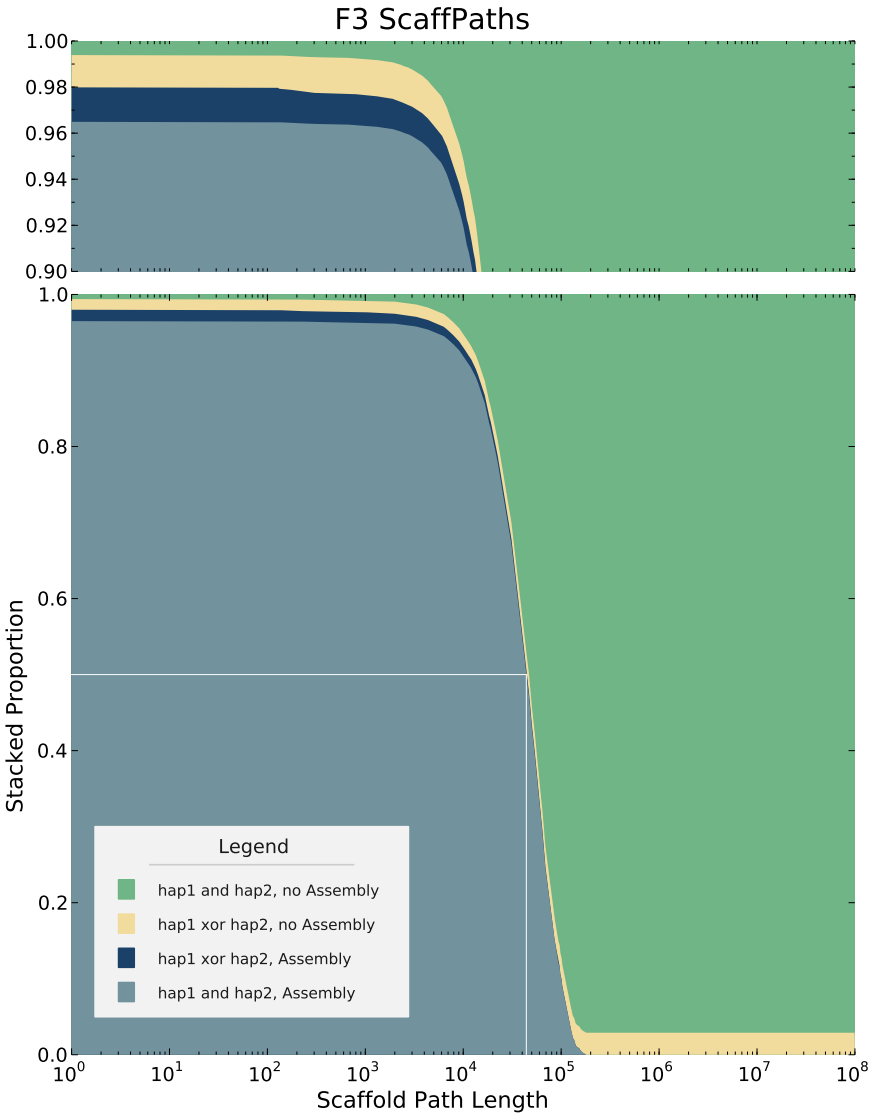


Figure 3.76: F3 scaffolds caption goes here.

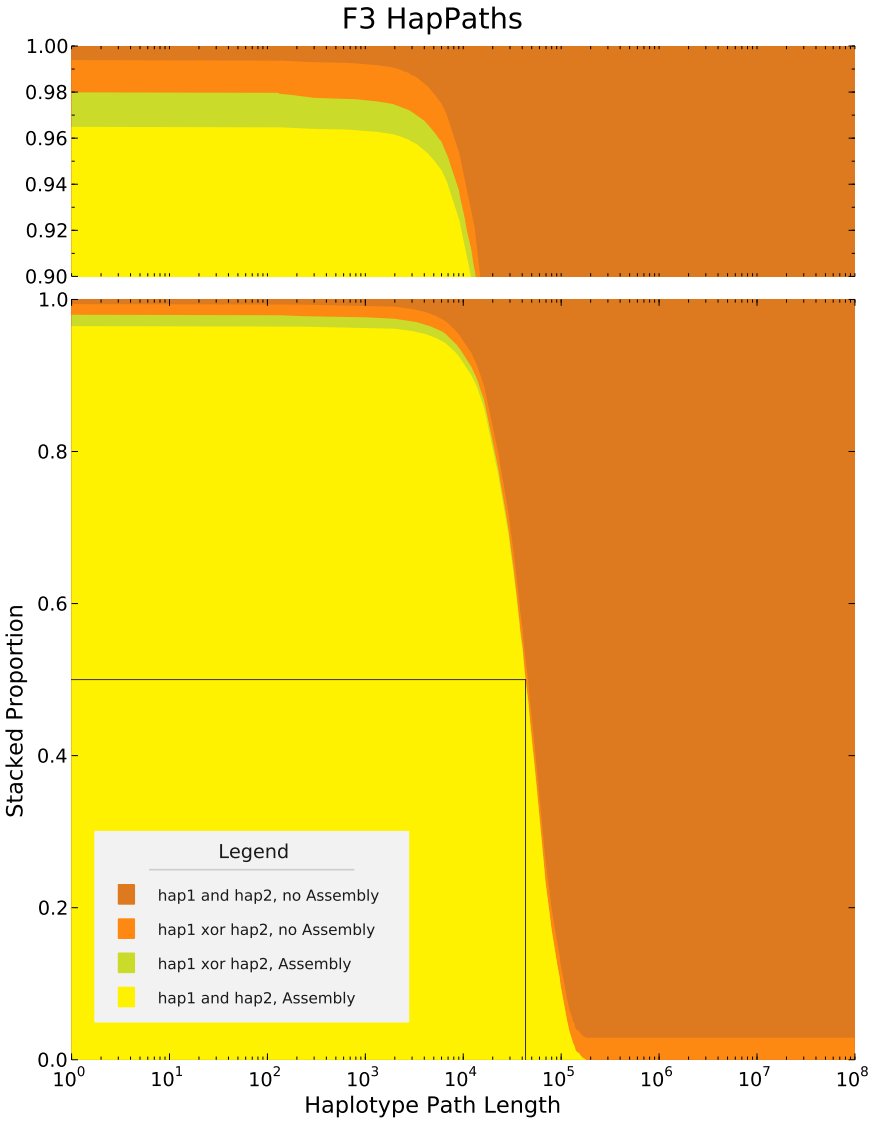


Figure 3.77: F3 hapPaths caption goes here.

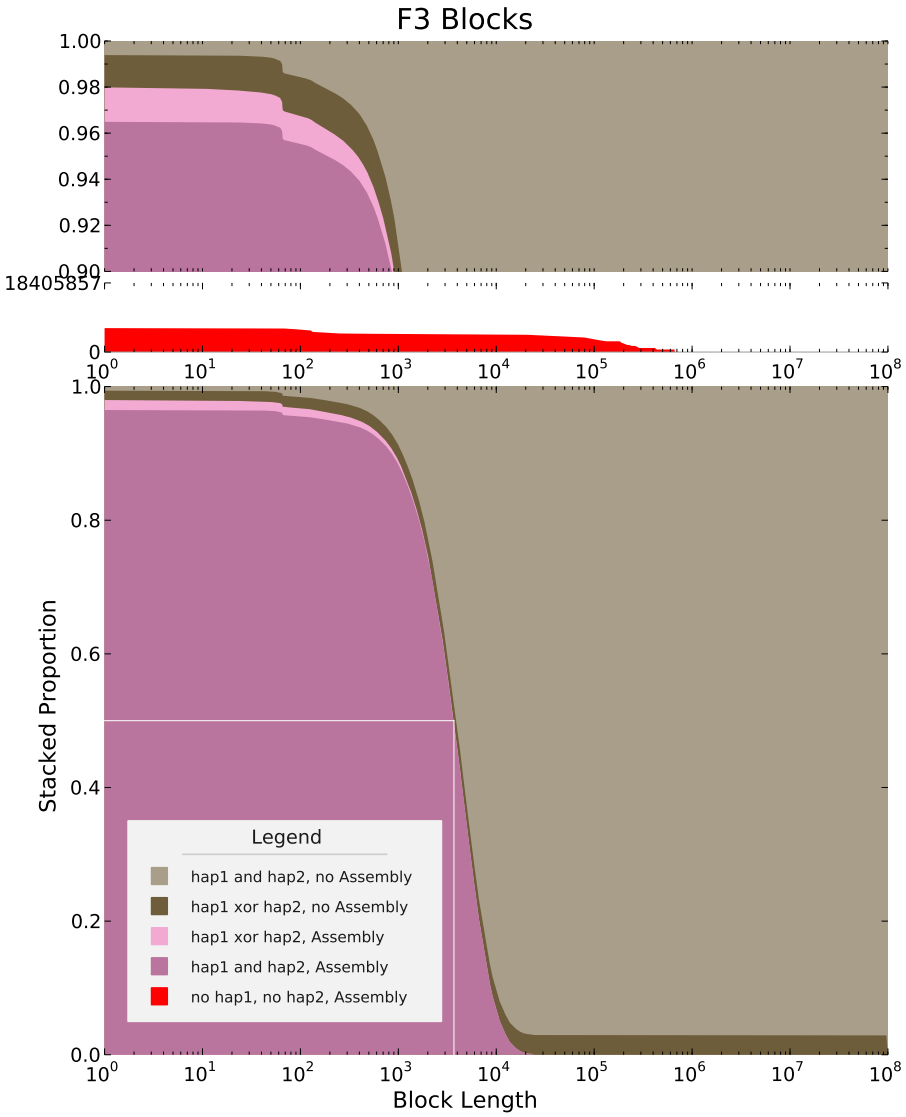


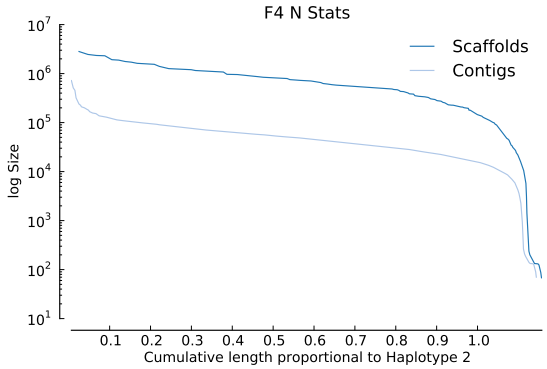
Figure 3.78: F3 blocks caption goes here.

F4

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
M5	0.98667	0.98679	0.98655	0.99969
F4	0.98649	0.98675	0.98624	0.99928
F2	0.98644	0.98676	0.98611	0.99924

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	28,681	67	115.00	133	4,540.72	159.00	2,820,201	62,235.74	130,232,384
Contigs	32,134	67	120.00	133	4,009.66	182.00	721,724	15,571.24	128,846,436

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	100,042,257 – 100,613,611	99,368,372 – 99,897,240	198,735,383.0 – 199,790,857.0	78 – 186
Heterozygous	389,681 – 397,504	386,235 – 393,291	769,900.0 – 783,534.0	1,245 – 1,433
Indel	2,772,485 – 3,173,973	1,274,187 – 1,583,172	2,544,716.0 – 3,160,763.0	1,716 – 1,892

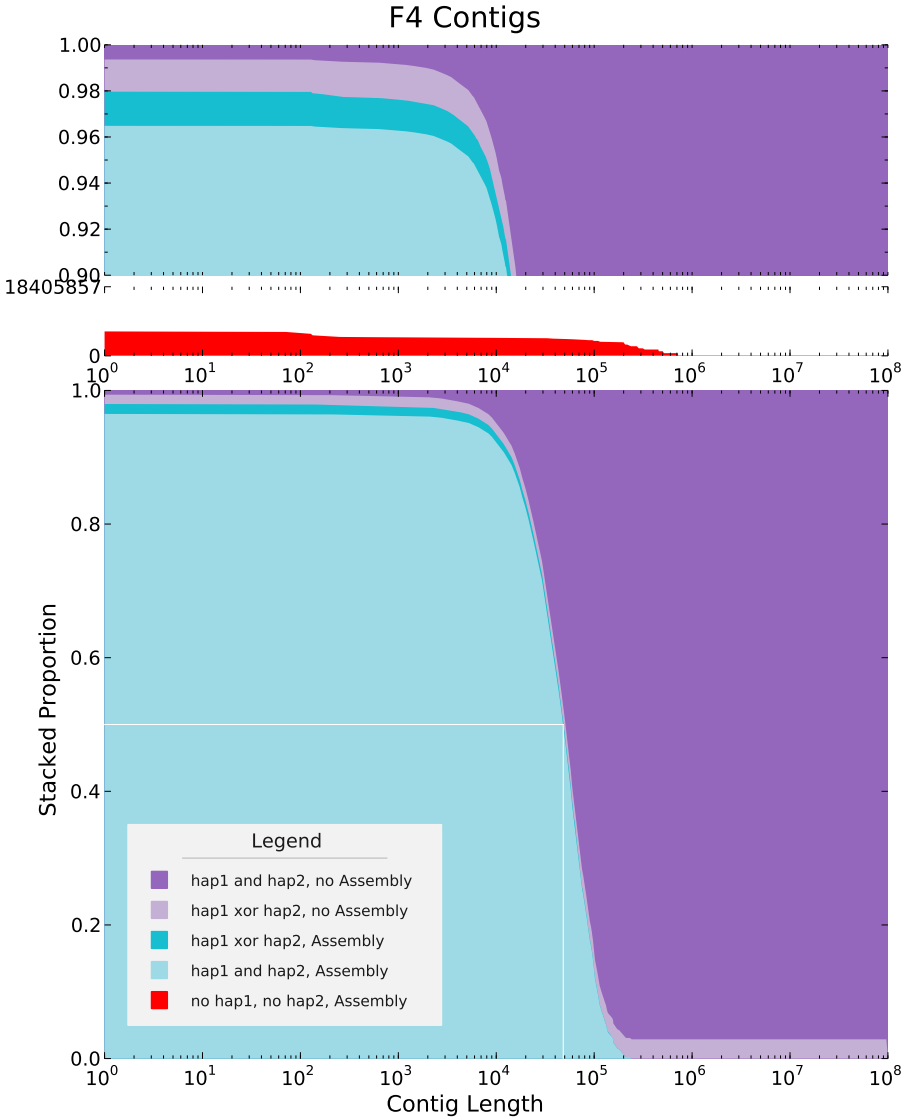


Figure 3.79: F4 contigs caption goes here.

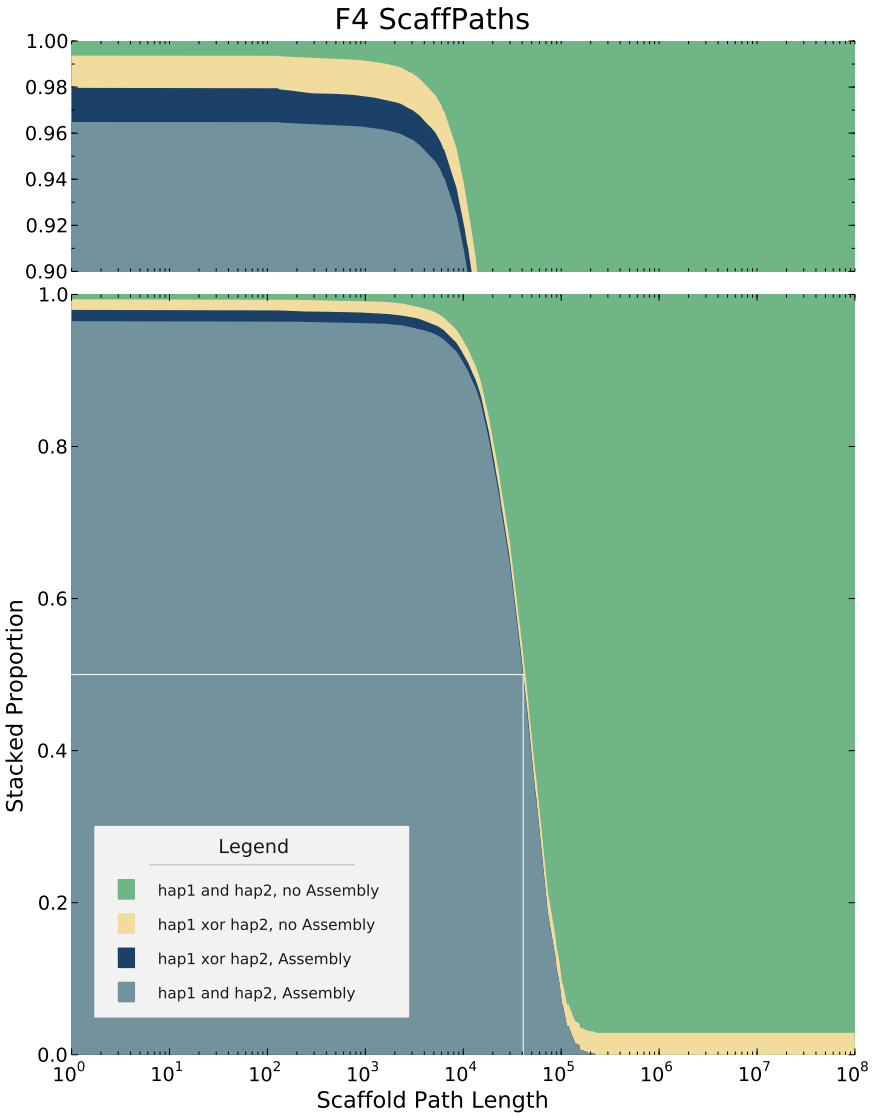


Figure 3.80: F4 scaffolds caption goes here.

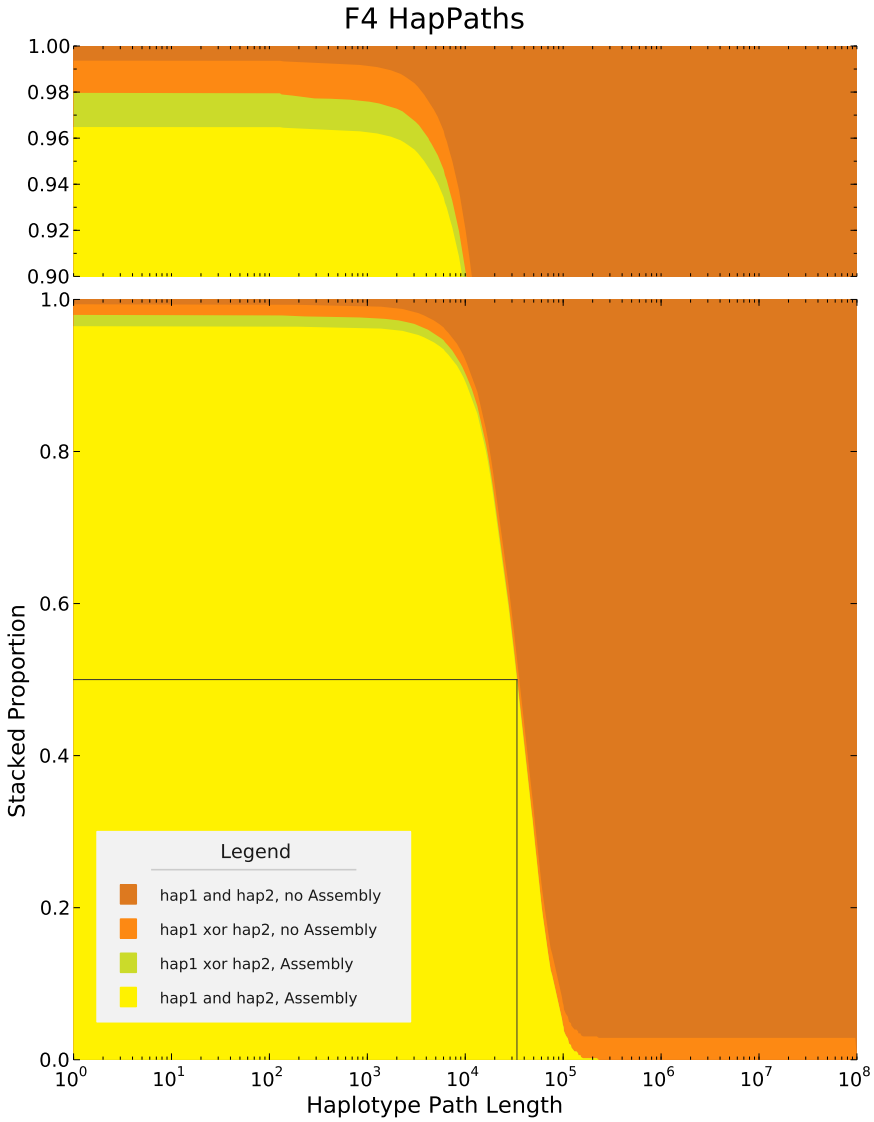


Figure 3.81: F4 hapPaths caption goes here.

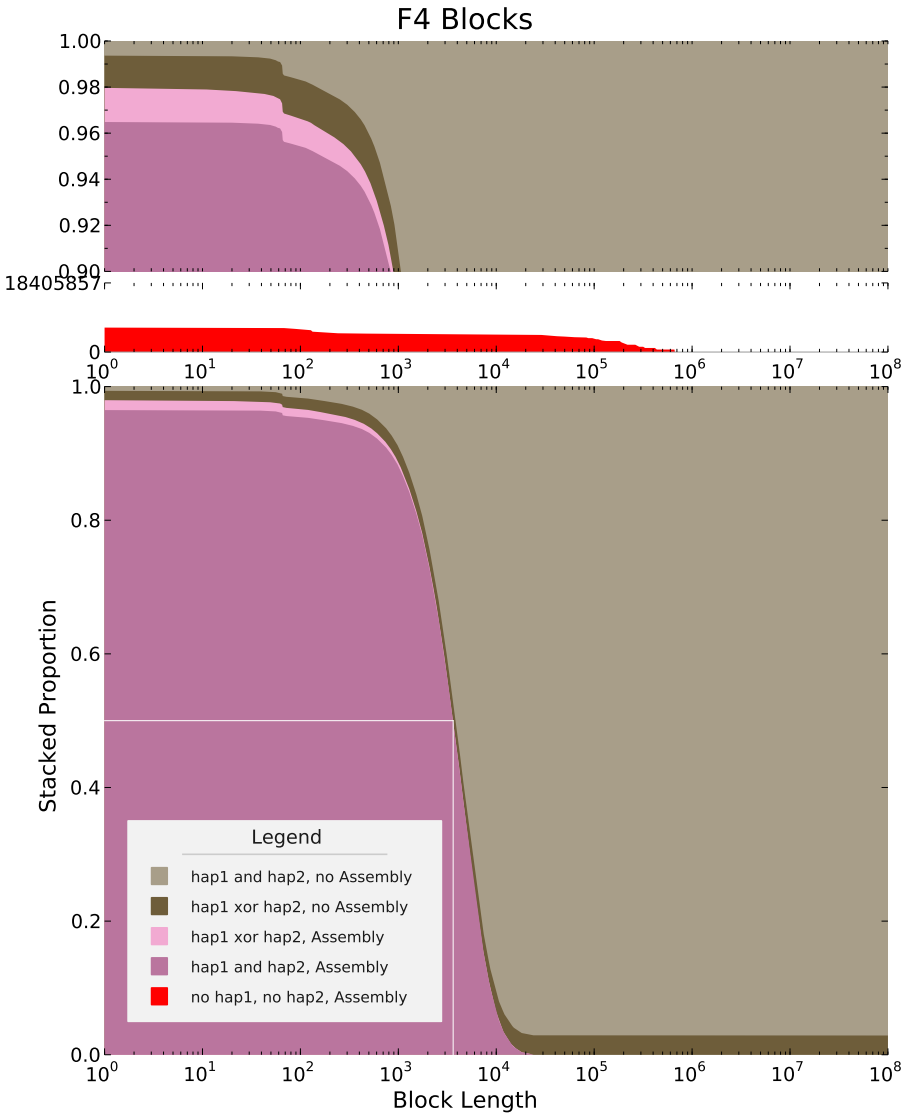


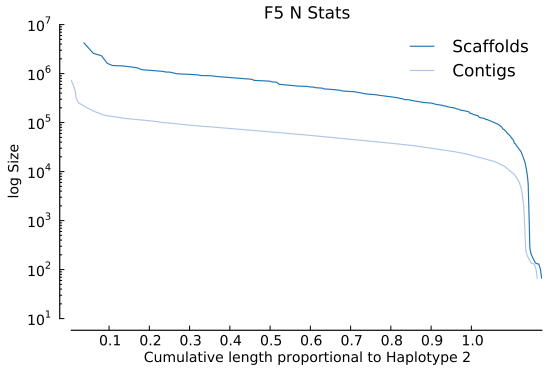
Figure 3.82: F4 blocks caption goes here.

F5

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
B1	0.98694	0.98719	0.98668	0.99790
F5	0.98691	0.98727	0.98653	0.99934
M3	0.98674	0.98685	0.98662	0.99977

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	25,812	67	113.00	133	5,122.60	161.00	4,232,700	62,574.52	132,224,524
Contigs	28,683	67	117.00	133	4,566.80	183.00	721,871	17,726.51	130,989,538

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	97,818,962 – 98,365,958	97,192,029 – 97,699,187	194,382,777.0 – 195,395,532.0	76 – 155
Heterozygous	381,627 – 389,289	378,382 – 385,345	753,520.0 – 766,820.0	1,580 – 1,837
Indel	2,767,272 – 3,164,605	1,261,291 – 1,566,319	2,518,988.0 – 3,127,222.0	1,735 – 1,934

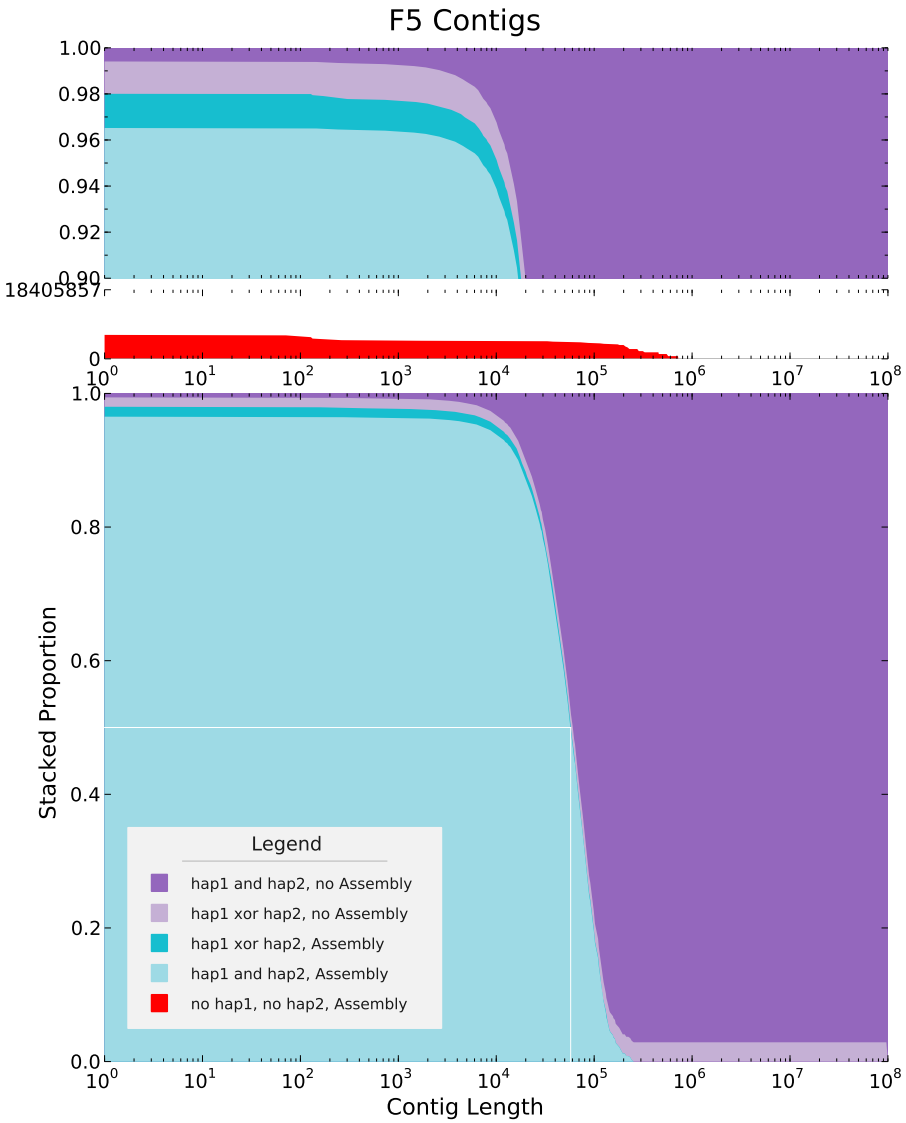


Figure 3.83: F5 contigs caption goes here.

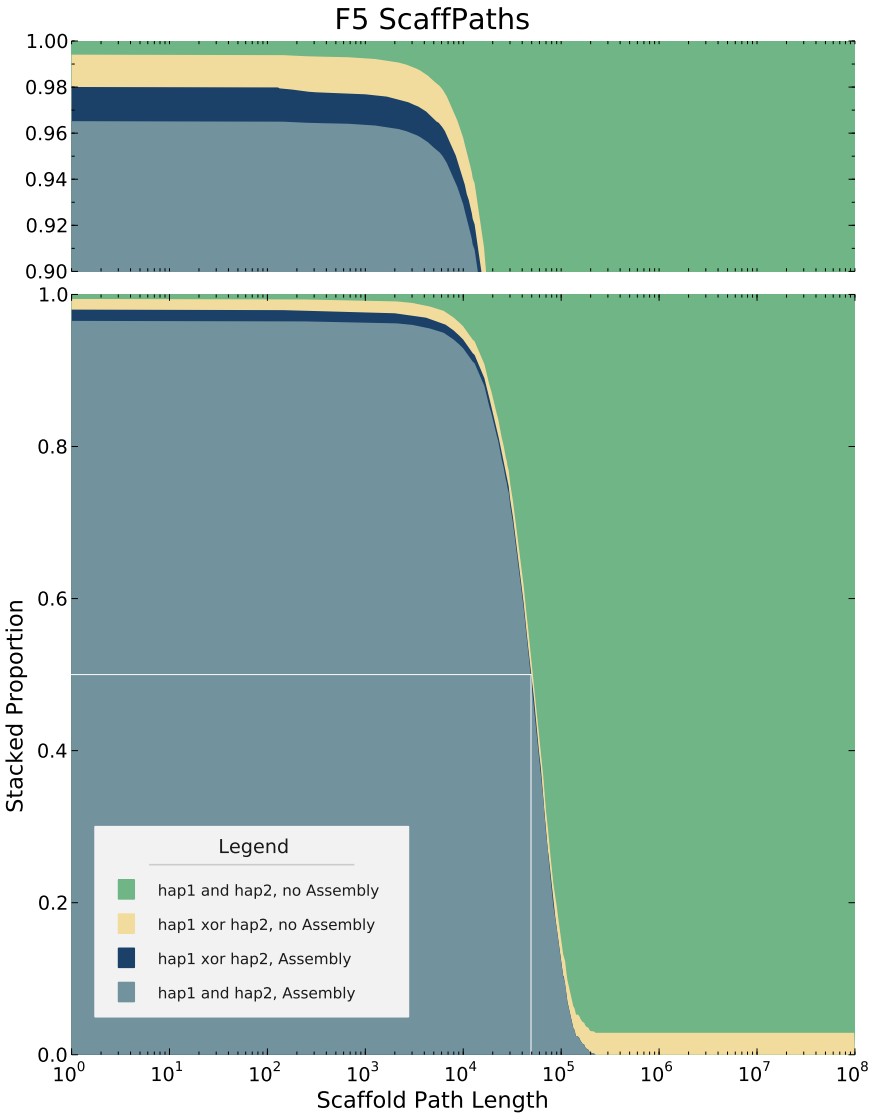


Figure 3.84: F5 scaffolds caption goes here.

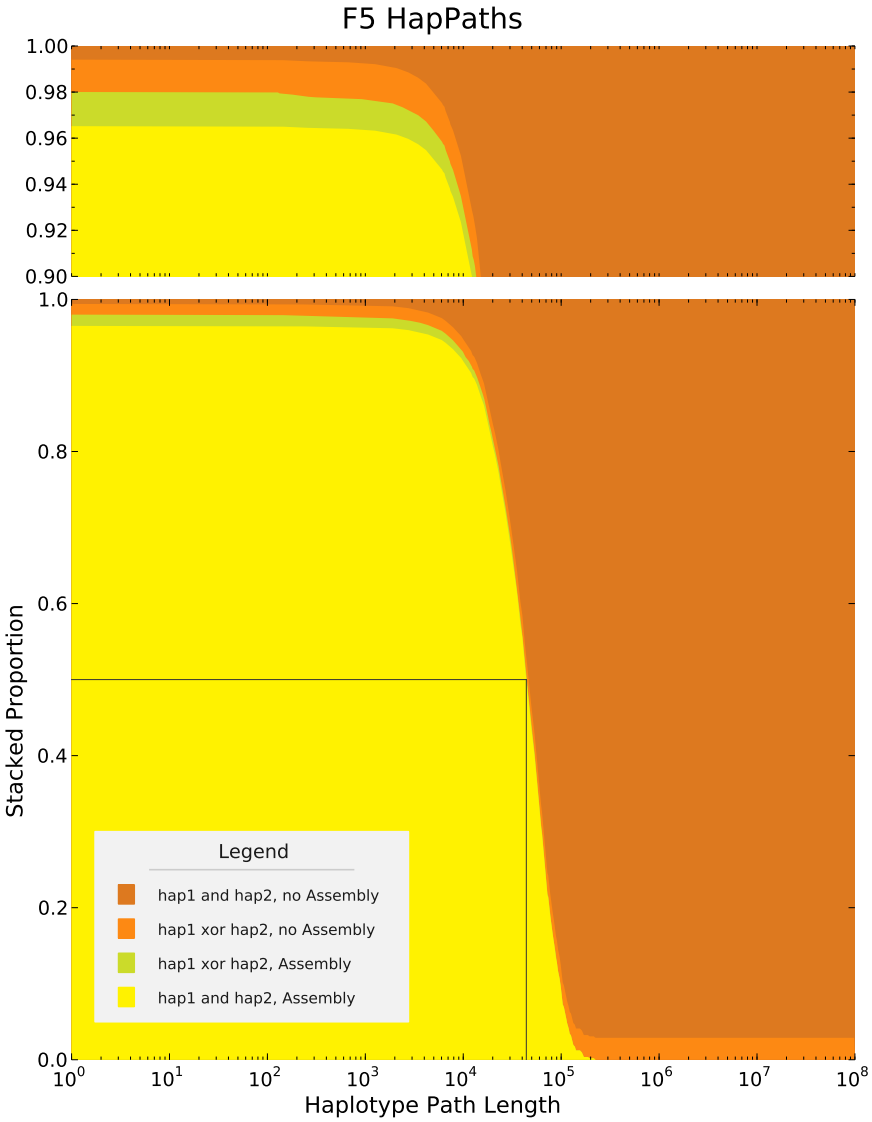


Figure 3.85: F5 hapPaths caption goes here.

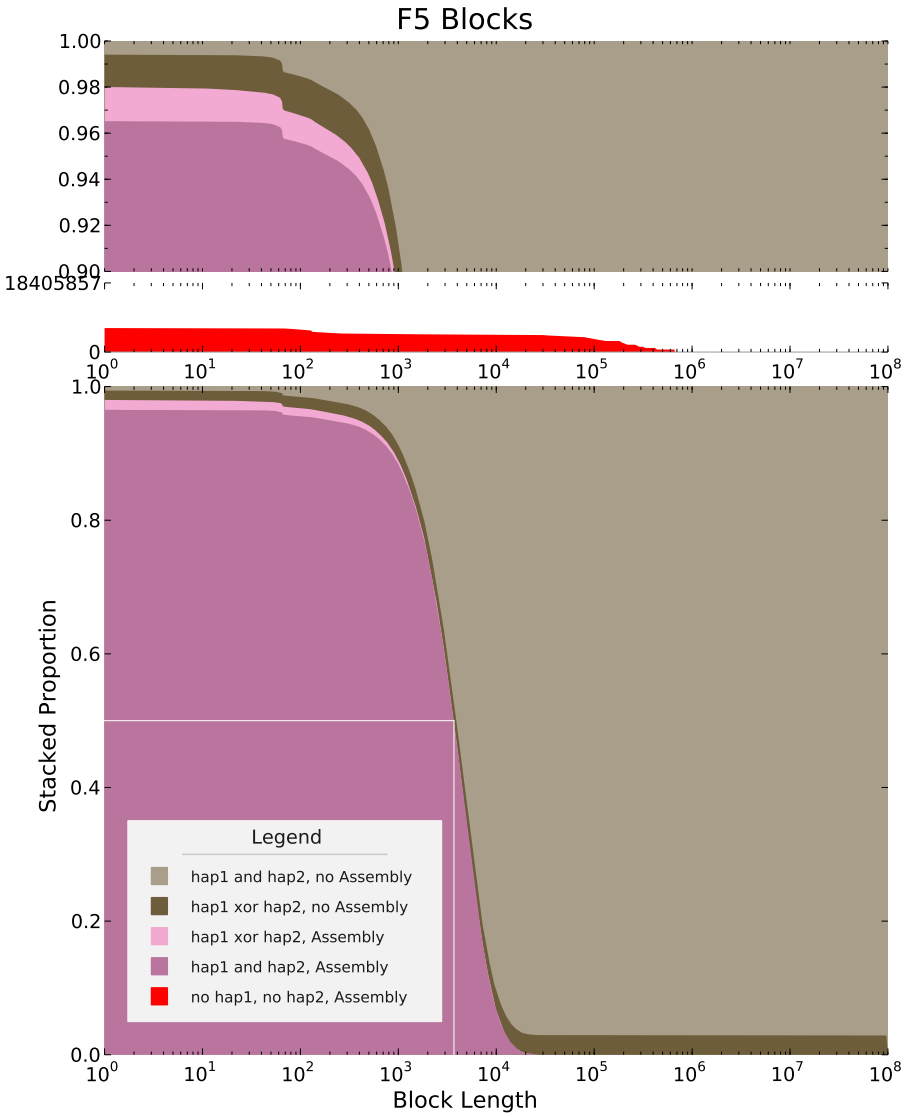


Figure 3.86: F5 blocks caption goes here.

3.2.7 G, Plant Genome Assembly Group

Affiliation: DOE Joint Genome Insititute, USA
Contact: Jarrod Chapman
Software: **Meraculous**
Number of entries: 1

ID	Total	Hap 1	Hap 2	Bac
G1	0.96922	0.96951	0.96894	0.99498

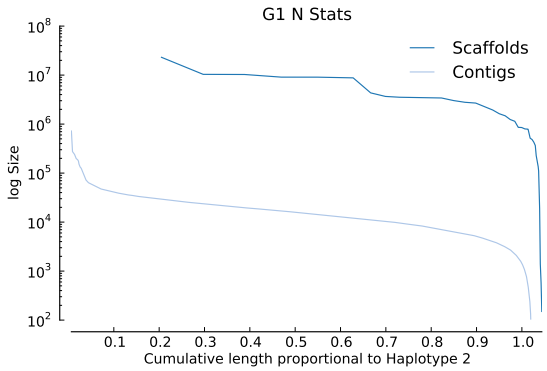
Assemblies:

G1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
W7	0.96984	0.97006	0.96961	0.99806
G1	0.96922	0.96951	0.96894	0.99498
W6	0.96892	0.96918	0.96865	0.99764

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	966	150	217.00	332	121,614.06	799.00	23,079,946	1,066,938.72	117,479,181
Contigs	14,817	100	999.00	4,452	7,744.12	10,738.00	722,469	12,207.13	114,744,669

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,156,013 – 110,346,770	106,984,207 – 107,681,757	213,955,608.0 – 215,310,228.0	41 – 64
Heterozygous	413,051 – 441,091	401,549 – 411,073	800,976.0 – 812,732.0	0 – 0
Indel	1,933,605 – 2,304,180	813,626 – 970,831	1,624,880.0 – 1,928,602.0	1,043 – 1,098

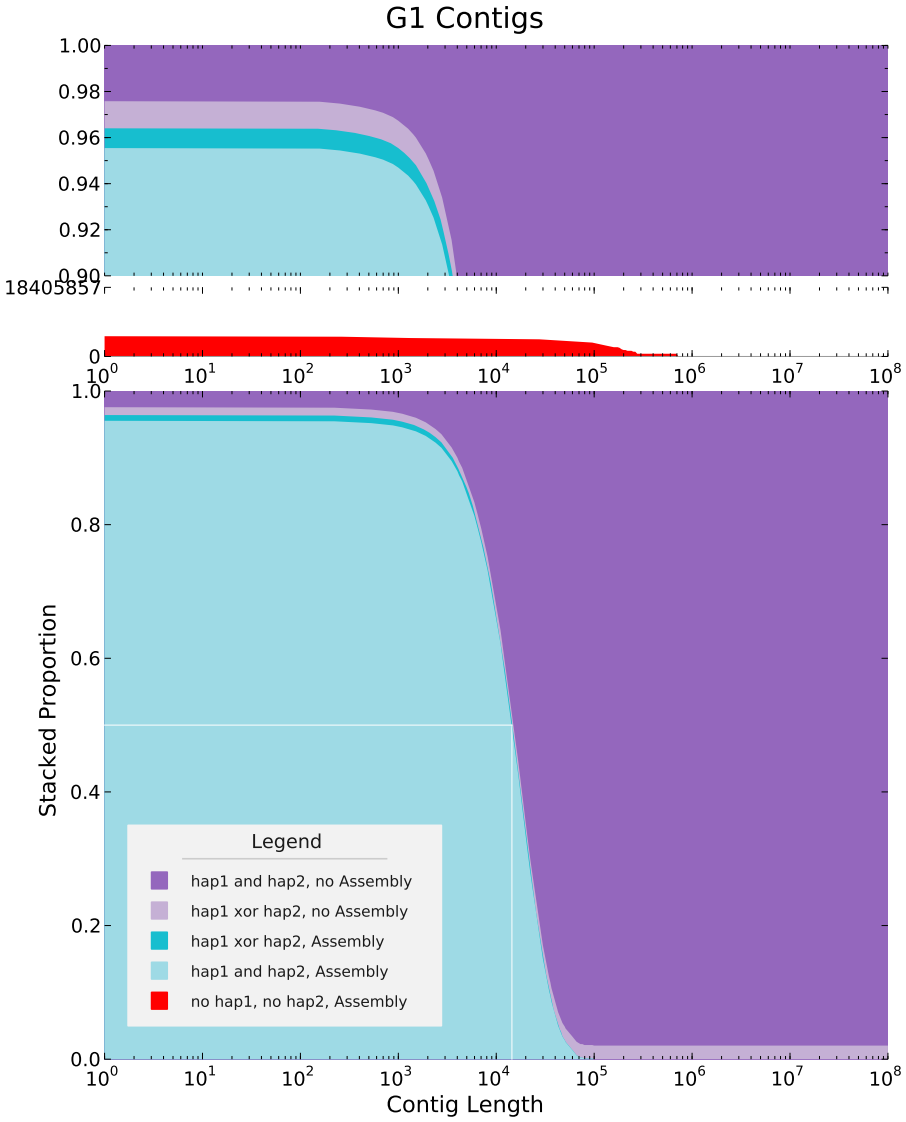


Figure 3.87: G1 contigs caption goes here.

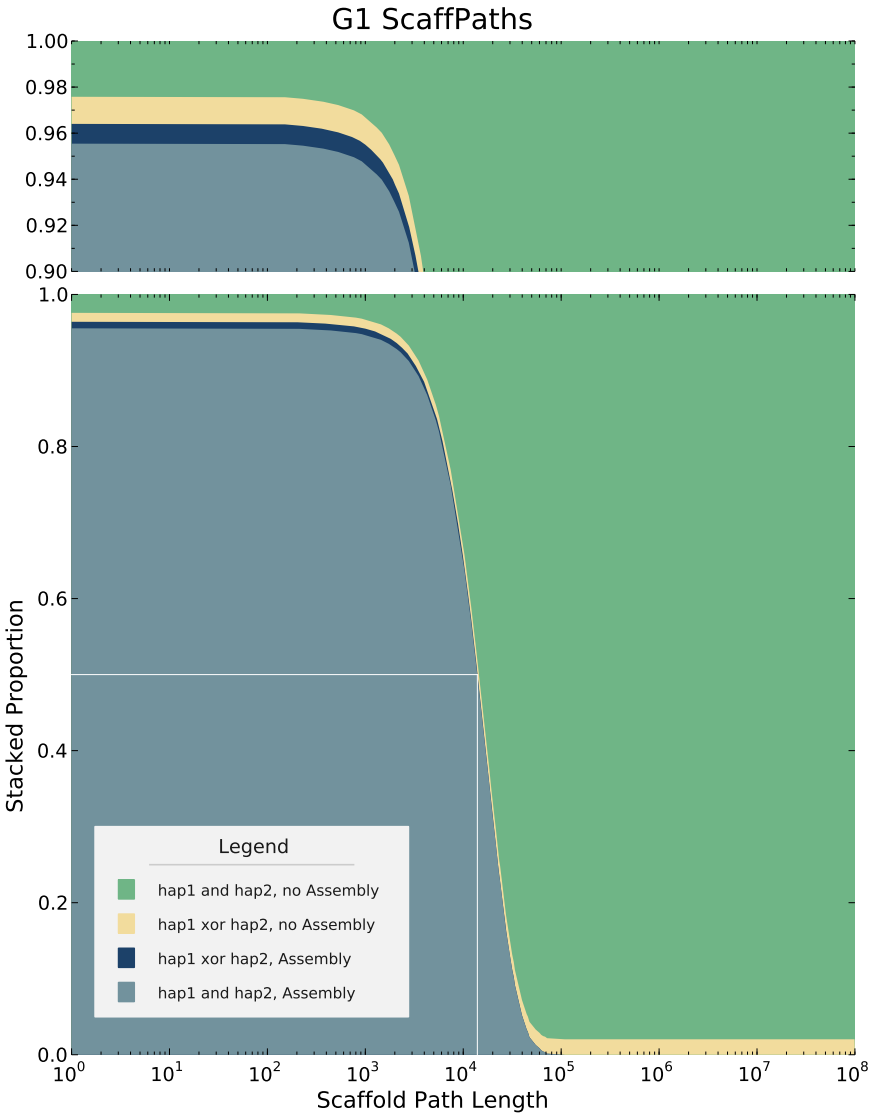


Figure 3.88: G1 scaffolds caption goes here.

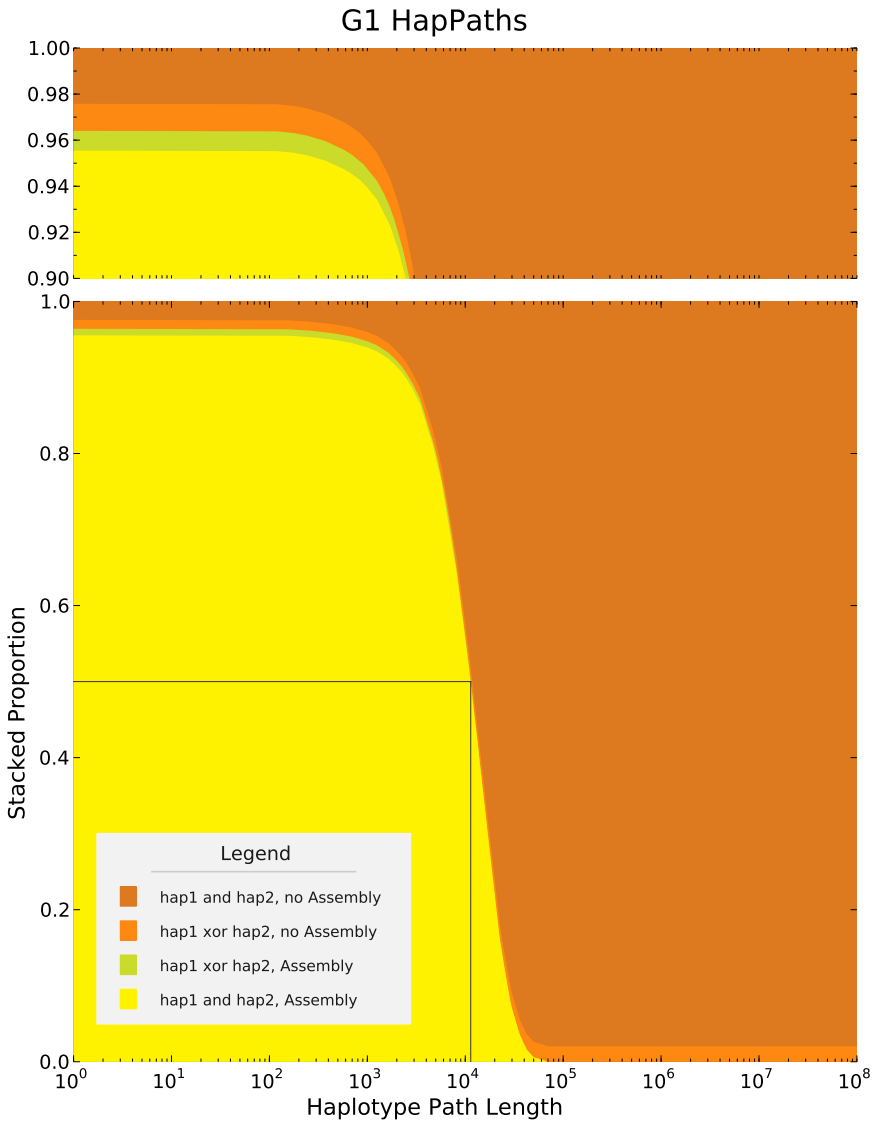


Figure 3.89: G1 hapPaths caption goes here.

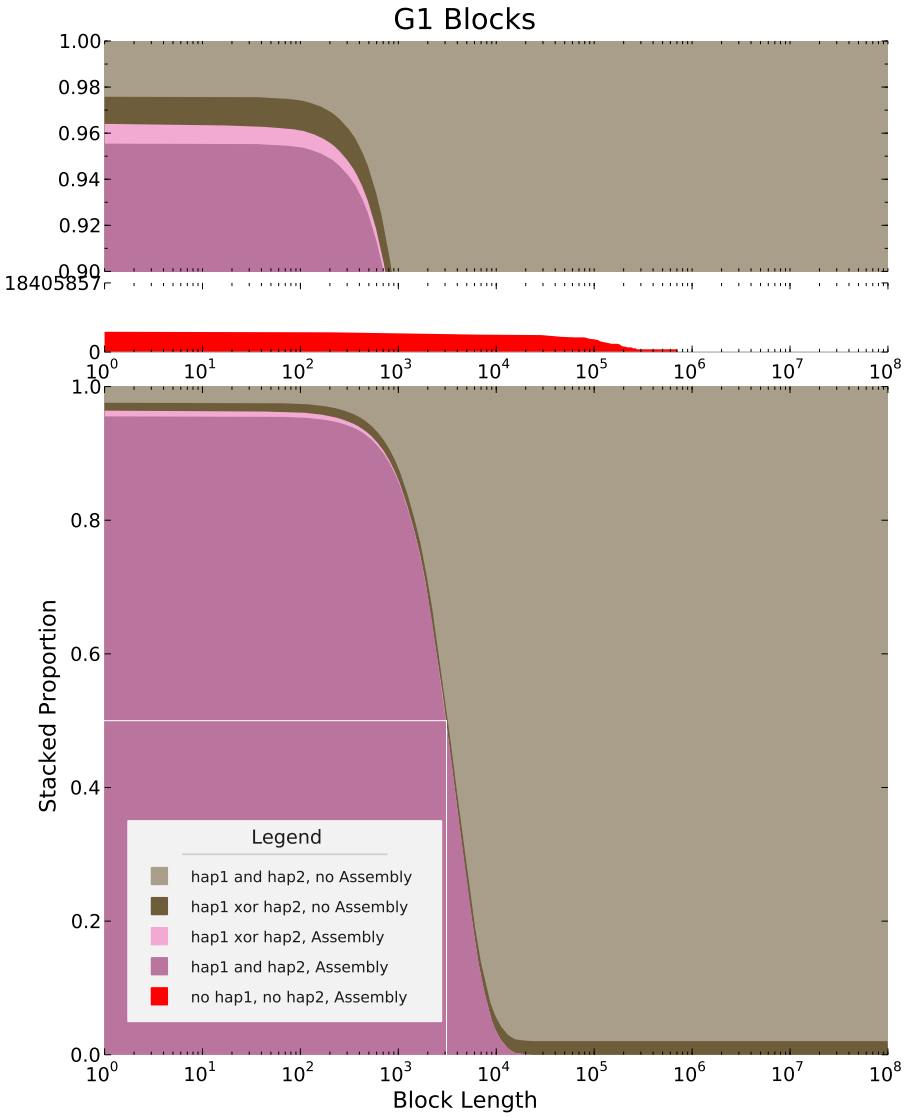


Figure 3.90: G1 blocks caption goes here.

3.2.8 H, Team Symbiose

Affiliation: L'IRISA (Institut de recherche en informatique et systèmes aléatoires), France
Contact: Rayan Chikhi
Software: **Monument**
Number of entries: 5

ID	Total	Hap 1	Hap 2	Bac
H4	0.95589	0.95589	0.95587	0.99681
H5	0.94518	0.94532	0.94503	0.99789
H1	0.93573	0.93582	0.93567	0.99709
H3	0.92715	0.92732	0.92696	0.99534
H2	0.92711	0.92728	0.92695	0.99536

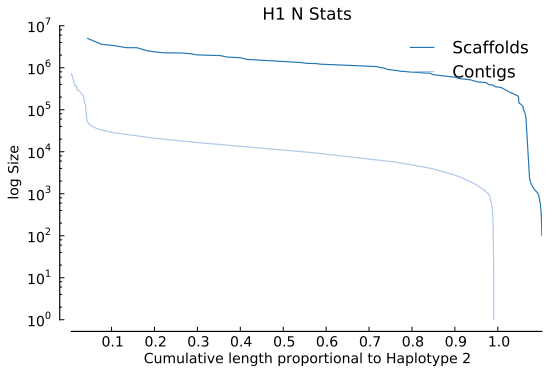
Assemblies:

H1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
J1	0.94234	0.94249	0.94219	0.99517
H1	0.93573	0.93582	0.93567	0.99709
H3	0.92715	0.92732	0.92696	0.99534

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	3,924	100	445.00	1,001	31,619.64	1,317.00	4,989,023	219,911.02	124,075,451
Contigs	19,446	1	1,210.00	3,222	5,732.46	7,715.00	722,170	11,067.75	111,473,322

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,583,668 – 109,825,235	102,511,065 – 103,290,776	205,011,450.0 – 206,559,850.0	4,571 – 6,496
Heterozygous	424,863 – 438,424	396,066 – 404,568	791,946.0 – 808,564.0	40 – 58
Indel	1,659,874 – 2,033,050	701,976 – 862,178	1,401,520.0 – 1,719,398.0	1,211 – 2,305

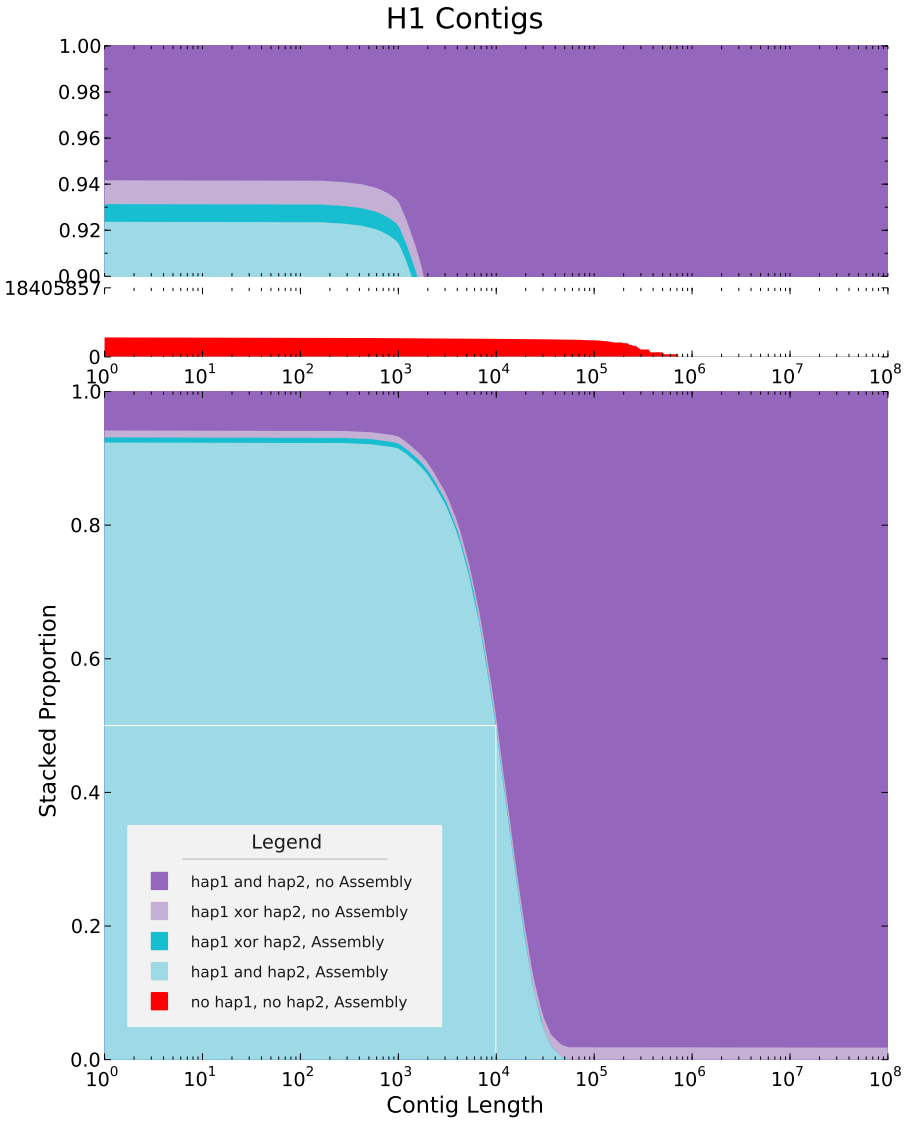


Figure 3.91: H1 contigs caption goes here.

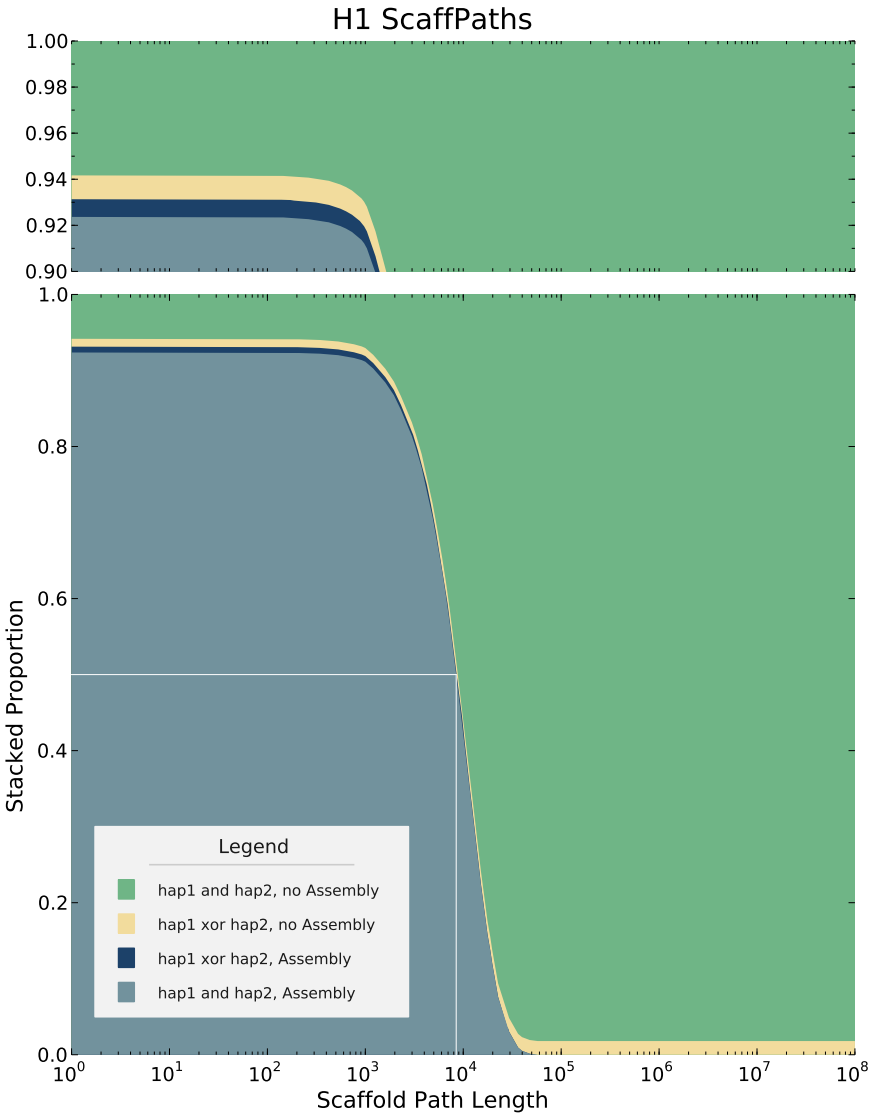


Figure 3.92: H1 scaffolds caption goes here.

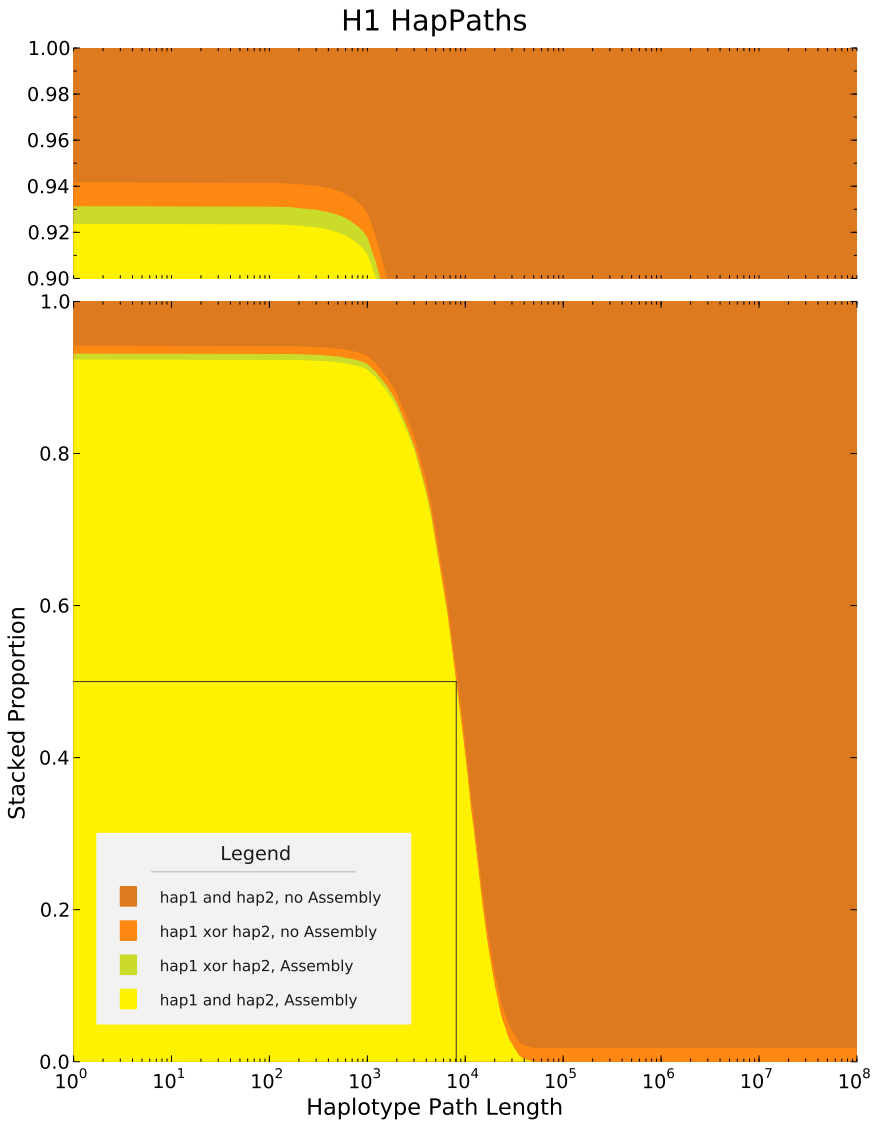


Figure 3.93: H1 hapPaths caption goes here.

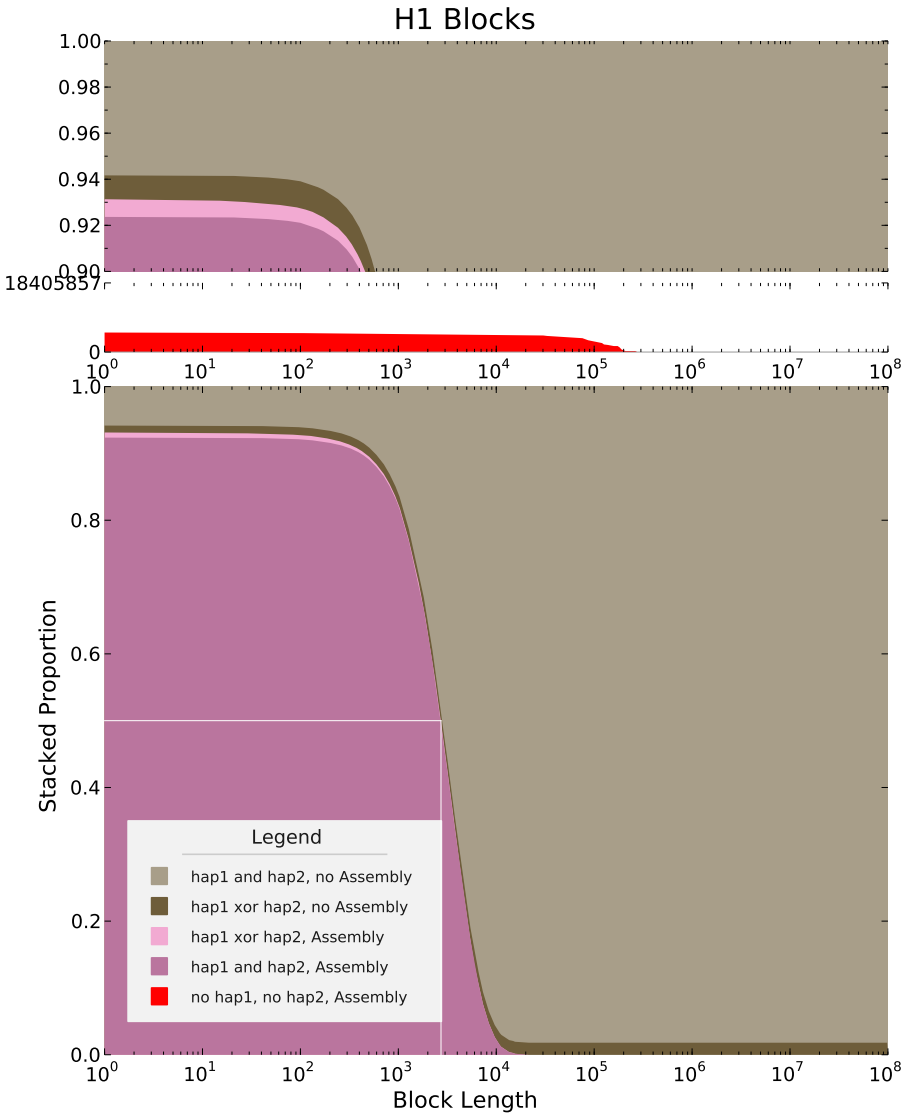


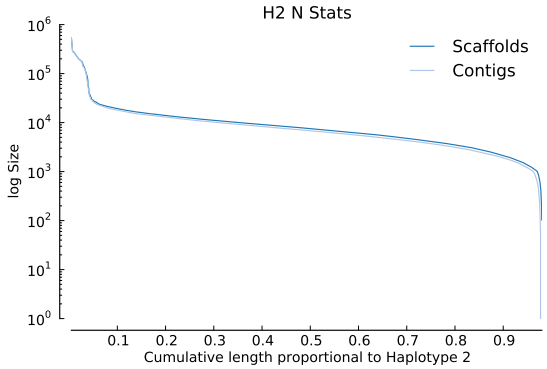
Figure 3.94: H1 blocks caption goes here.

H2

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
H3	0.92715	0.92732	0.92696	0.99534
H2	0.92711	0.92728	0.92695	0.99536
C2	0.91842	0.91878	0.91805	0.00000

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	22,181	100	1,714.00	3,417	4,971.83	6,512.00	541,160	7,954.59	110,280,191
Contigs	25,981	1	1,324.00	2,838	4,233.40	5,610.00	541,160	7,327.56	109,987,924

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,909,616 – 110,376,813	102,013,172 – 102,880,000	204,015,082.0 – 205,732,812.0	3,718 – 5,444
Heterozygous	424,731 – 440,953	391,759 – 400,831	783,186.0 – 800,830.0	38 – 57
Indel	1,549,284 – 1,921,180	626,343 – 778,688	1,250,642.0 – 1,552,804.0	985 – 2,006

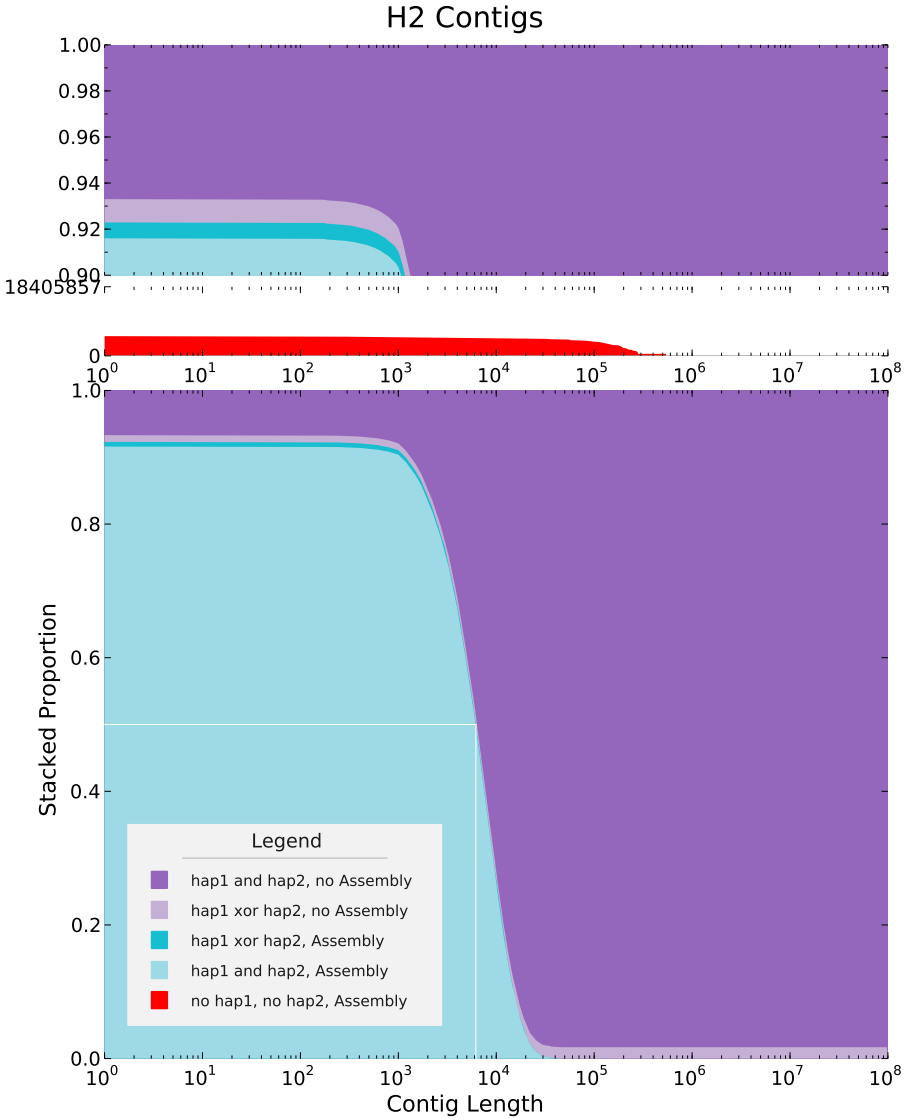


Figure 3.95: H2 contigs caption goes here.

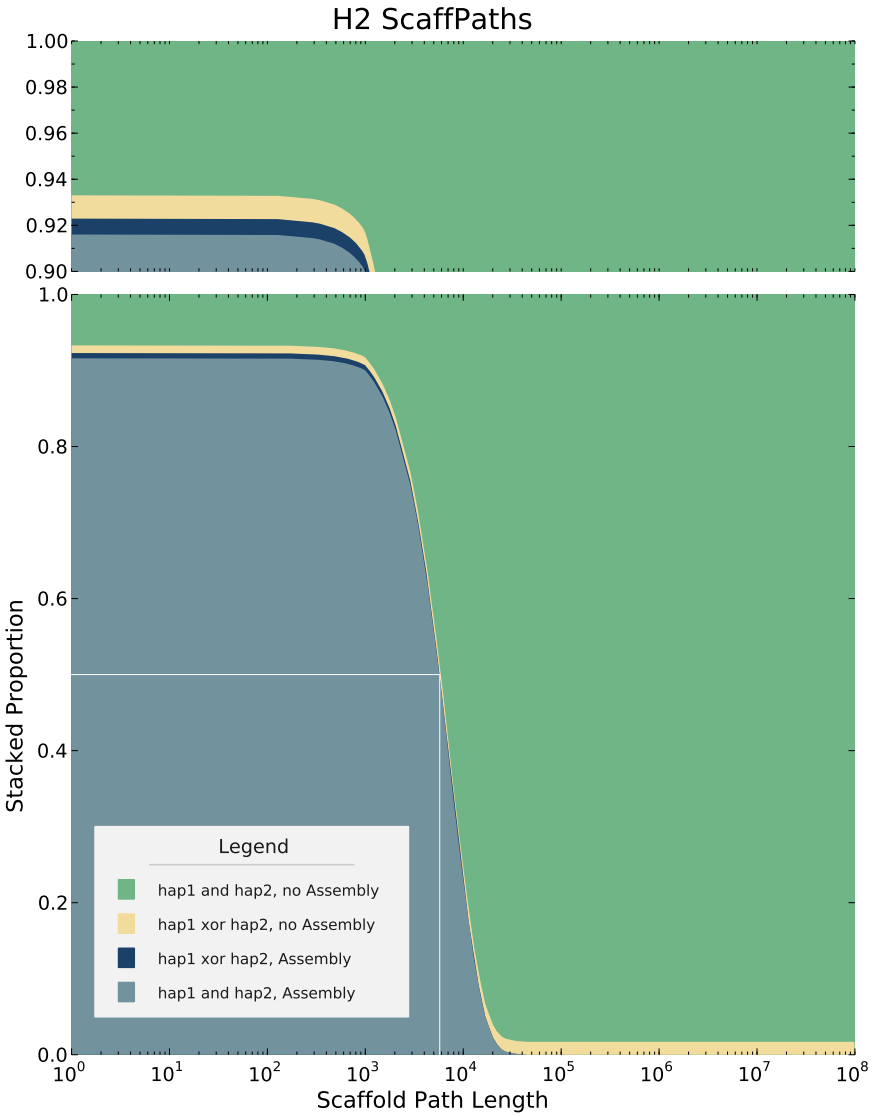


Figure 3.96: H2 scaffolds caption goes here.

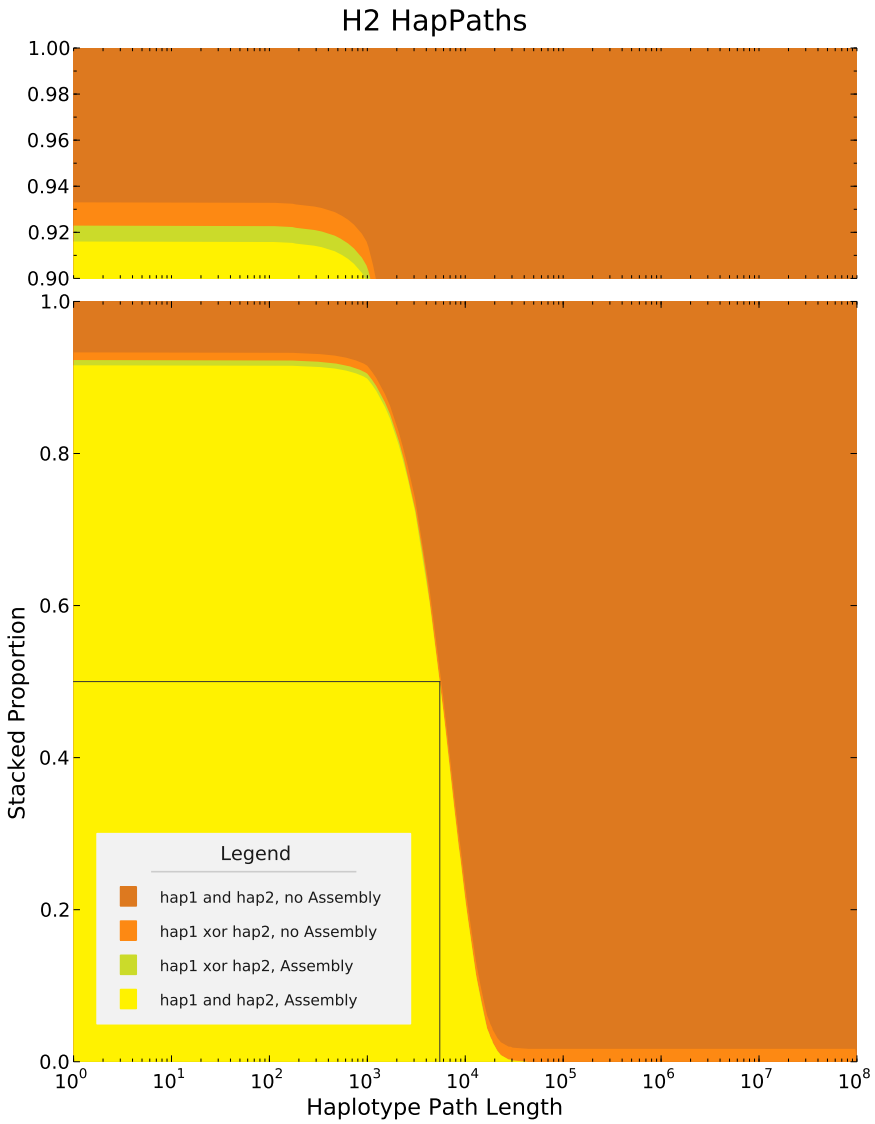


Figure 3.97: H2 hapPaths caption goes here.

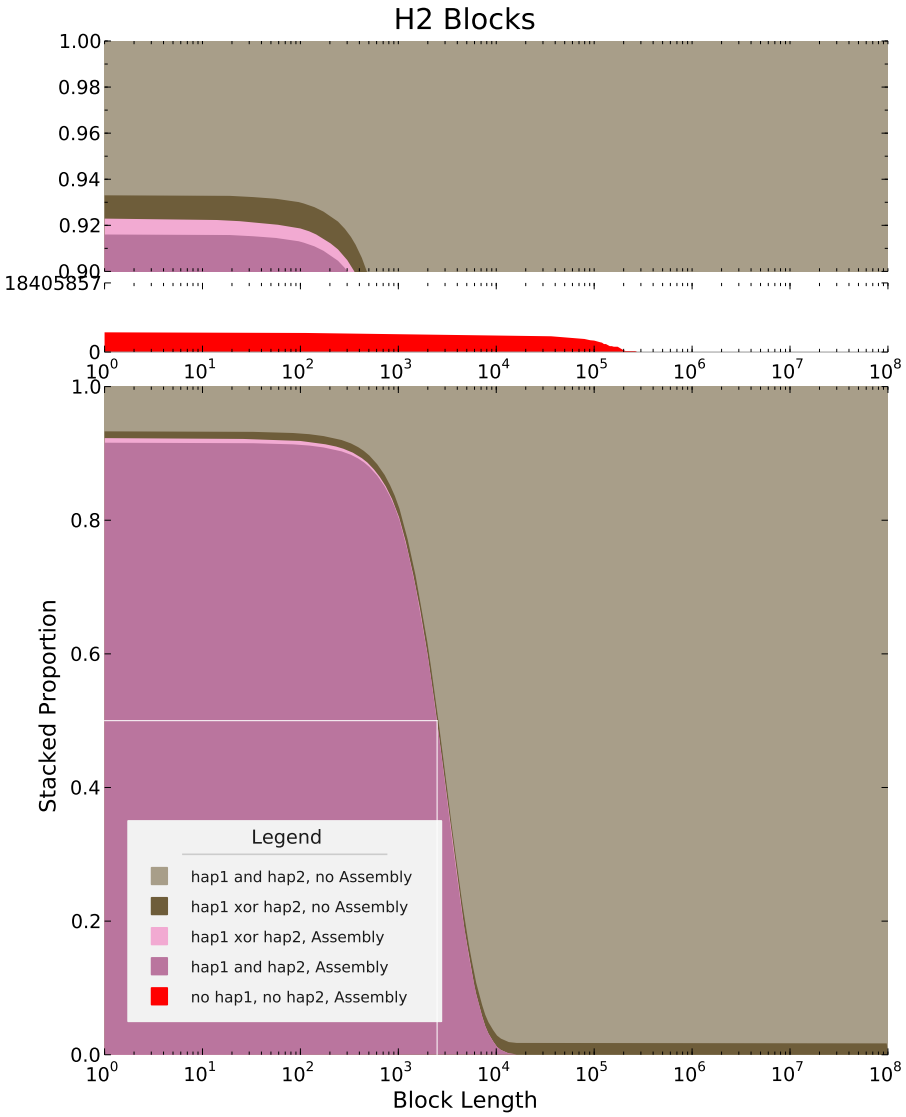


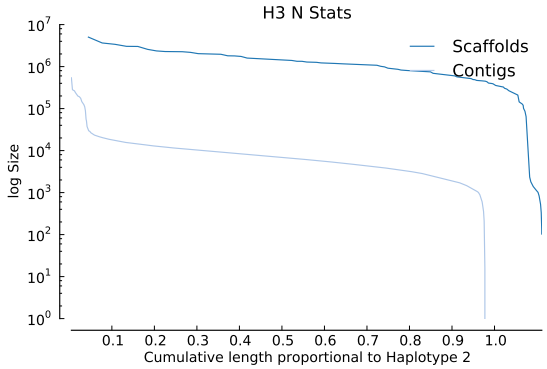
Figure 3.98: H2 blocks caption goes here.

H3

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
H1	0.93573	0.93582	0.93567	0.99709
H3	0.92715	0.92732	0.92696	0.99534
H2	0.92711	0.92728	0.92695	0.99536

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	3,924	100	445.00	1,001	31,868.19	1,317.00	5,037,481	221,724.26	125,050,796
Contigs	25,709	1	1,328.00	2,889	4,278.16	5,672.00	541,160	7,413.24	109,987,145

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,924,626 – 110,389,766	102,026,137 – 102,891,606	204,040,992.0 – 205,756,184.0	3,718 – 5,486
Heterozygous	424,776 – 440,926	391,811 – 400,821	783,282.0 – 800,804.0	41 – 57
Indel	1,530,034 – 1,901,797	617,914 – 770,538	1,233,808.0 – 1,536,504.0	981 – 2,008

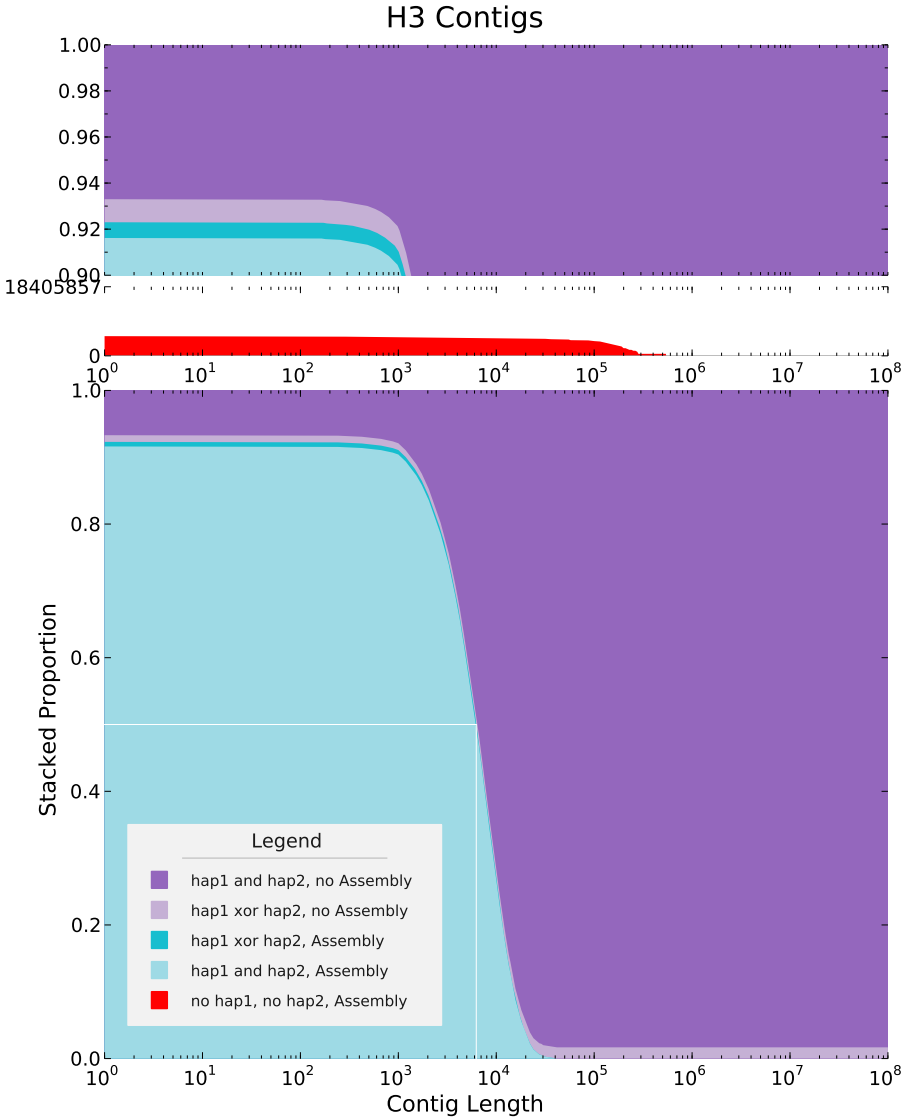


Figure 3.99: H3 contigs caption goes here.

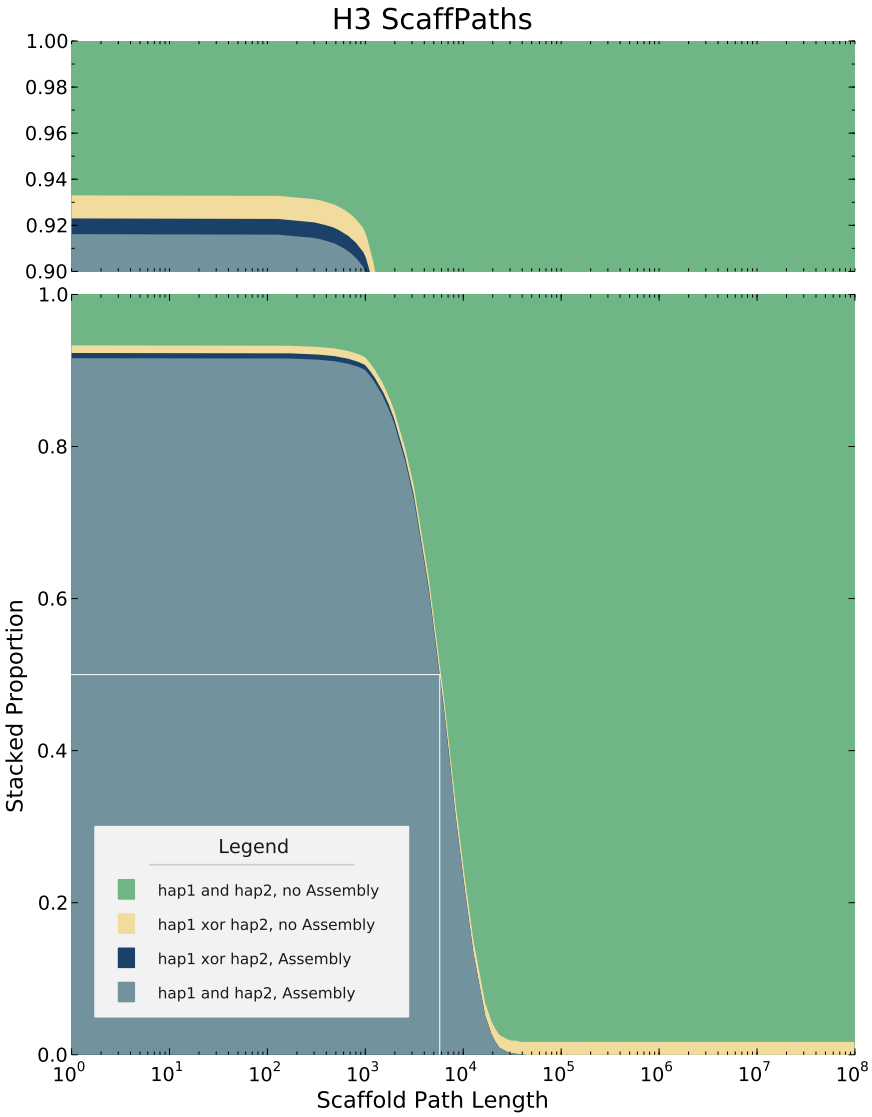


Figure 3.100: H3 scaffolds caption goes here.

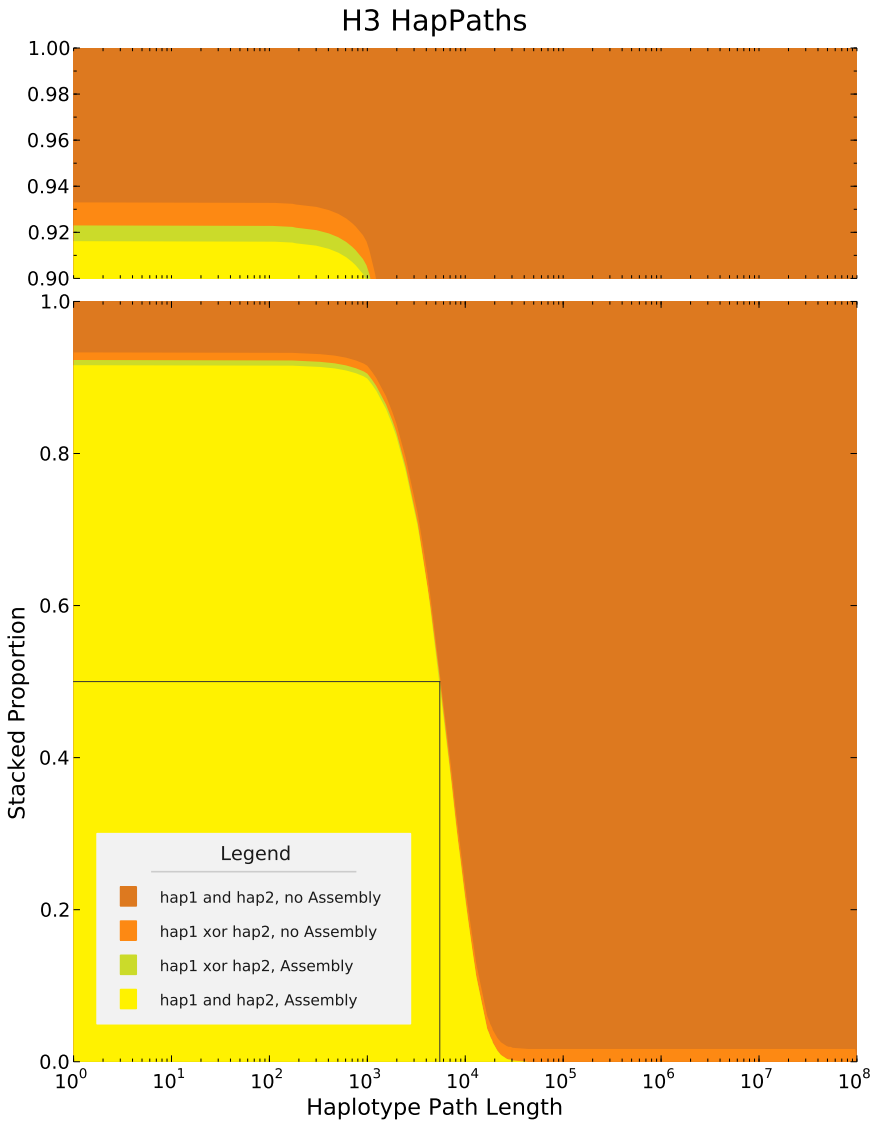


Figure 3.101: H3 hapPaths caption goes here.

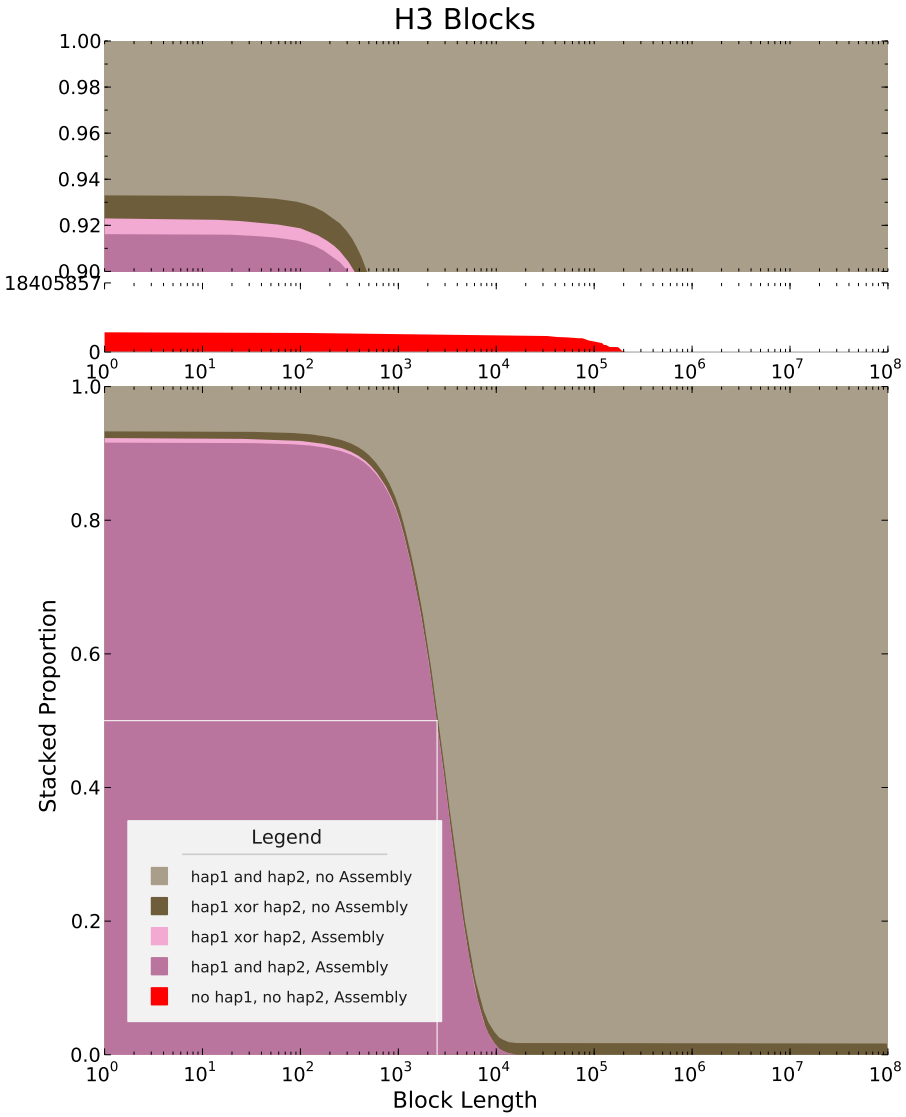


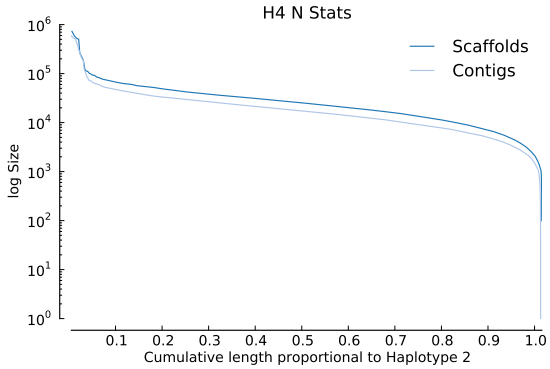
Figure 3.102: H3 blocks caption goes here.

H4

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
E2	0.95596	0.95640	0.95551	0.99215
H4	0.95589	0.95589	0.95587	0.99681
E3	0.95559	0.95601	0.95517	0.99317

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	8,517	100	3,153.00	7,812	13,418.66	17,636.00	726,869	19,995.28	114,286,718
Contigs	12,316	1	2,213.00	5,481	9,256.58	12,086.25	587,037	14,611.29	114,004,040

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,068,300 – 109,108,408	104,214,426 – 104,897,436	208,421,418.0 – 209,773,994.0	1,700 – 2,742
Heterozygous	424,958 – 435,871	407,517 – 415,375	814,802.0 – 830,070.0	17 – 25
Indel	2,161,999 – 2,535,782	949,061 – 1,123,141	1,895,060.0 – 2,240,472.0	1,449 – 2,301

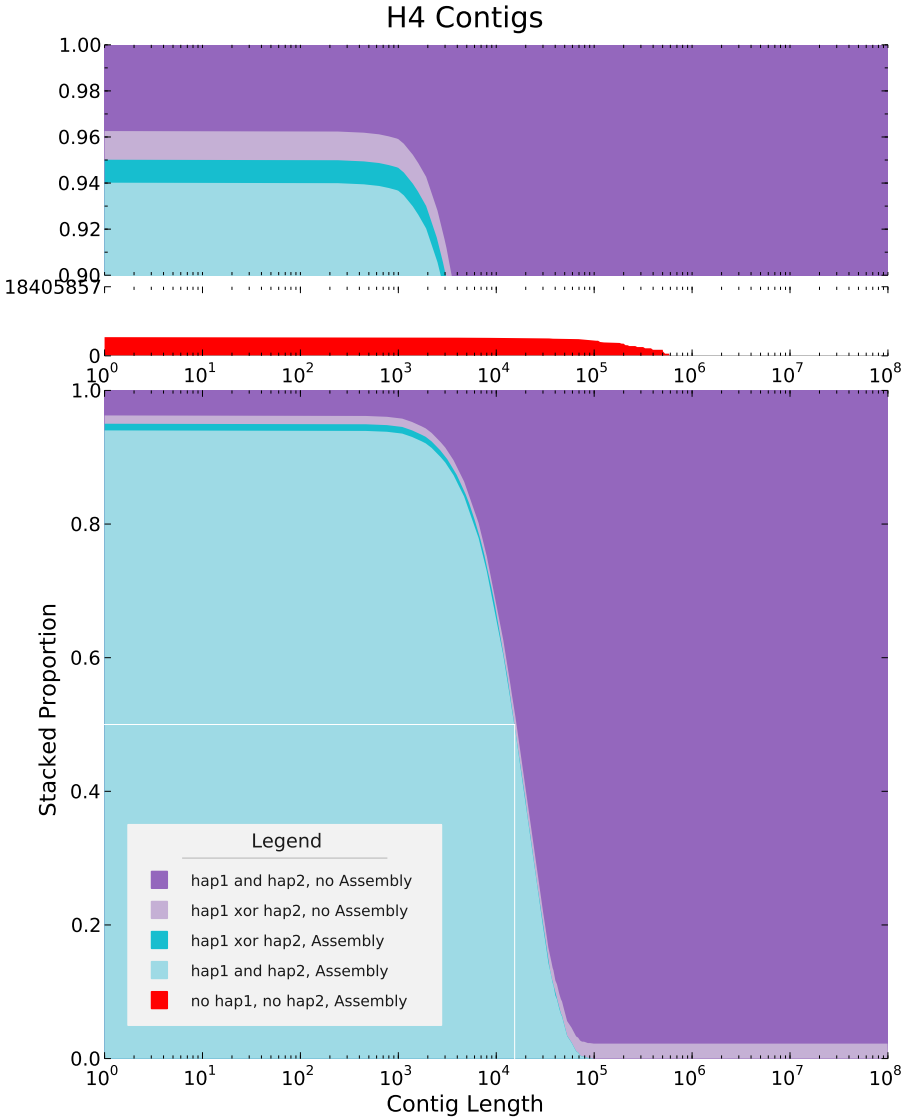


Figure 3.103: H4 contigs caption goes here.

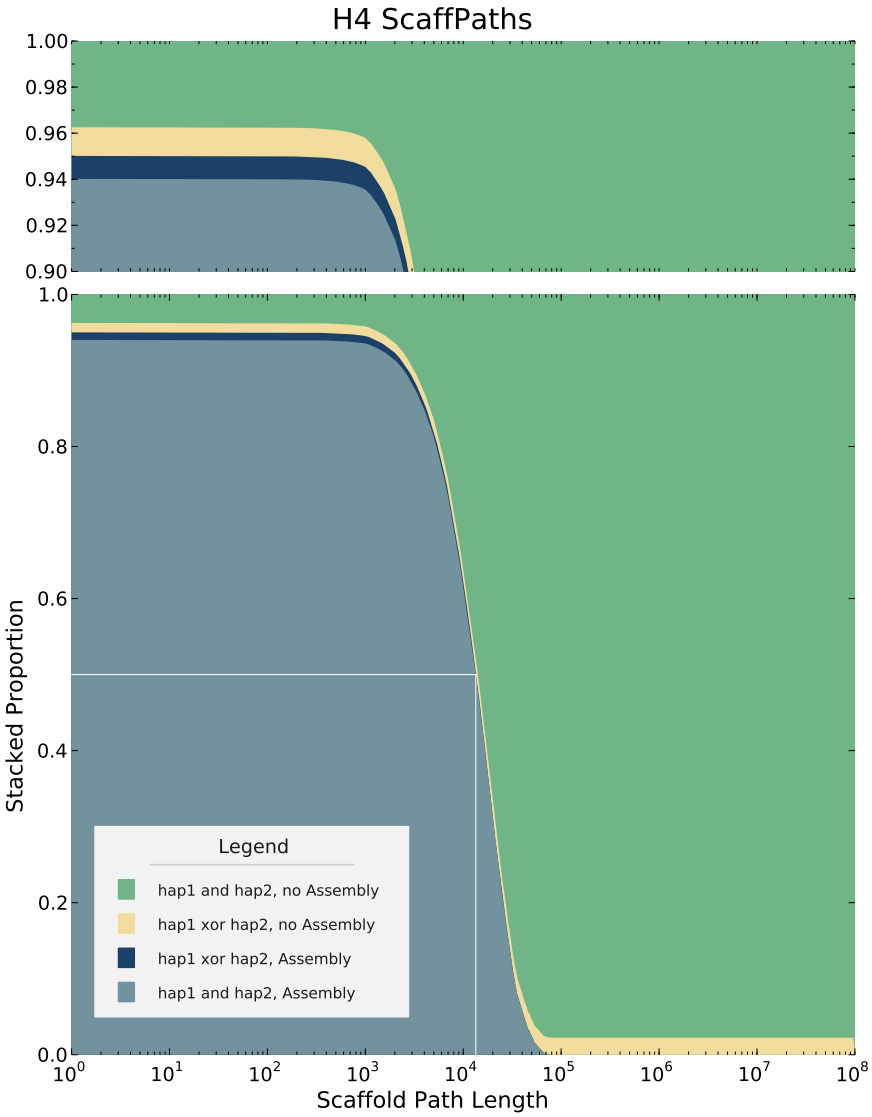


Figure 3.104: H4 scaffolds caption goes here.

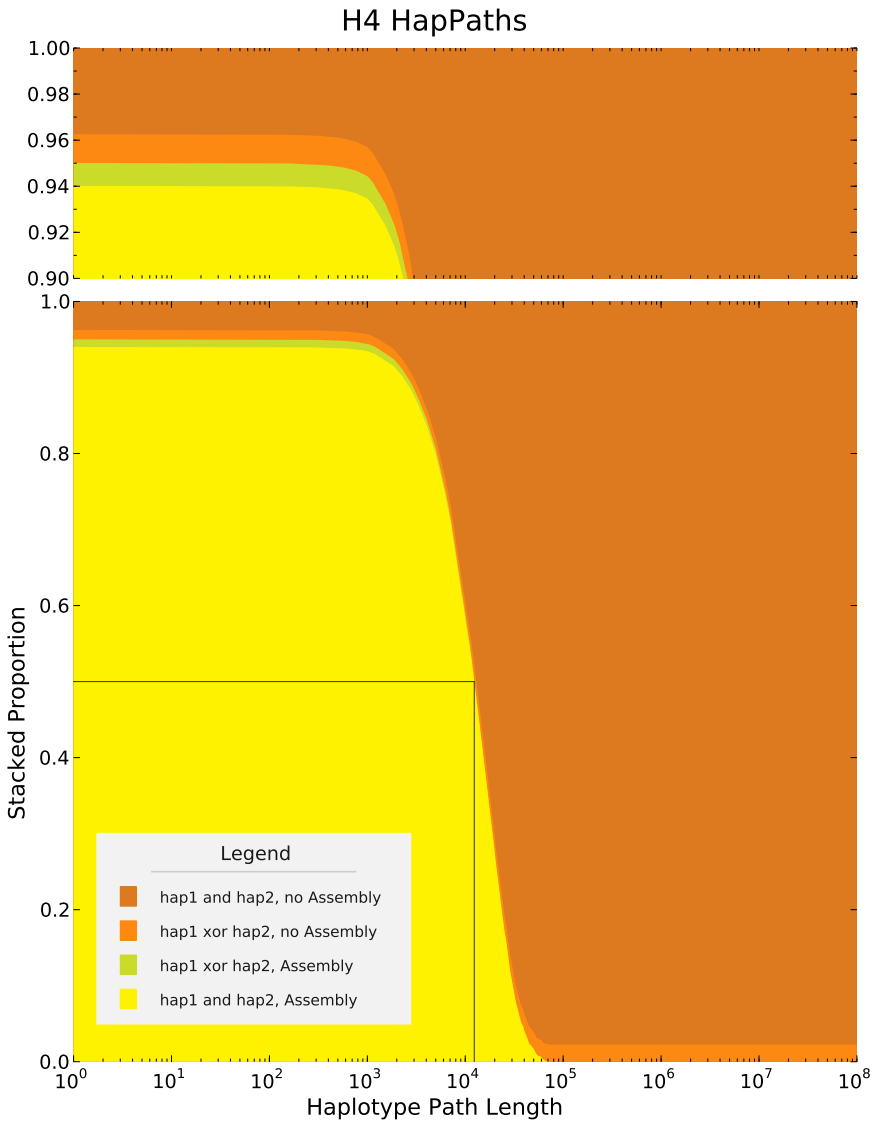


Figure 3.105: H4 hapPaths caption goes here.

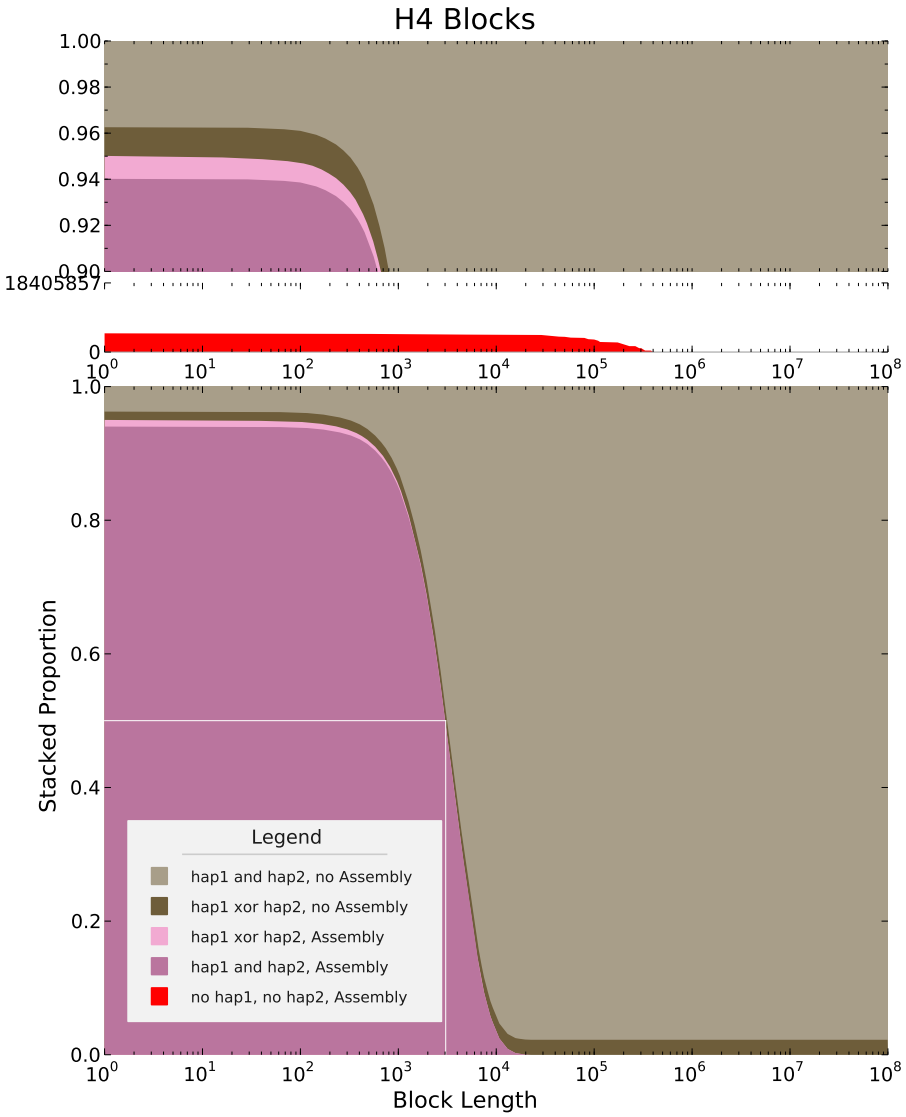


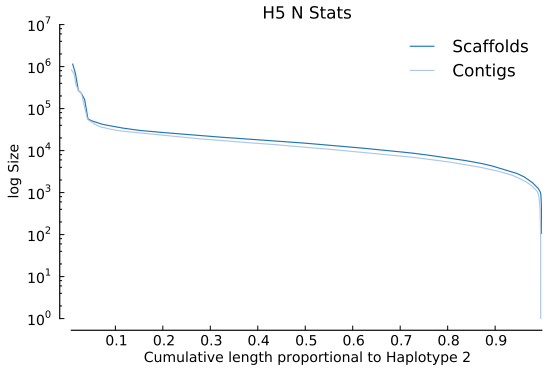
Figure 3.106: H4 blocks caption goes here.

H5

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
X3	0.95516	0.95535	0.95495	0.99560
H5	0.94518	0.94532	0.94503	0.99789
J1	0.94234	0.94249	0.94219	0.99517

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	12,904	100	2,484.75	5,525	8,705.94	11,666.50	1,143,876	15,681.50	112,341,484
Contigs	16,190	1	1,896.00	4,297	6,922.93	9,103.00	831,935	12,480.19	112,082,306

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,738,853 – 109,900,362	103,691,222 – 104,459,524	207,376,006.0 – 208,897,476.0	1,502 – 3,378
Heterozygous	427,347 – 439,318	404,349 – 412,998	808,458.0 – 825,312.0	24 – 40
Indel	1,997,393 – 2,372,708	887,453 – 1,060,393	1,772,218.0 – 2,114,642.0	1,335 – 2,704

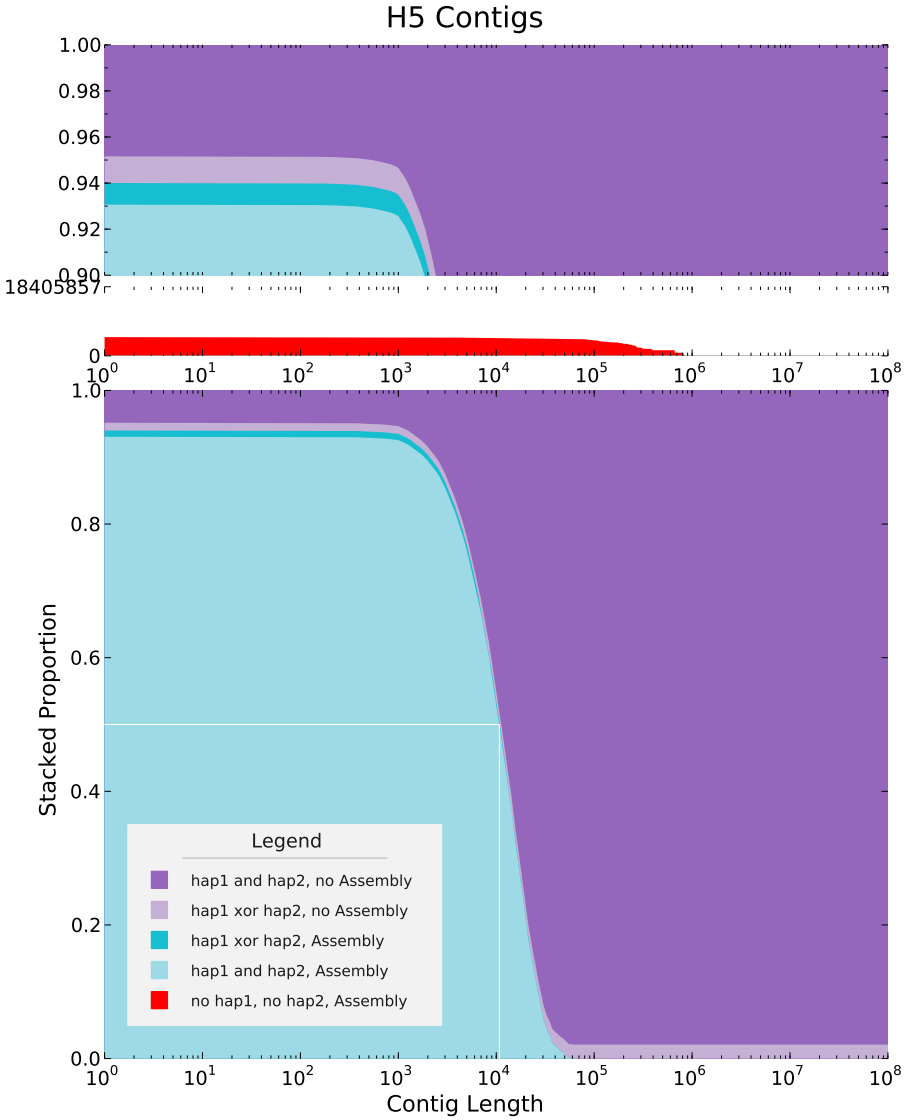


Figure 3.107: H5 contigs caption goes here.

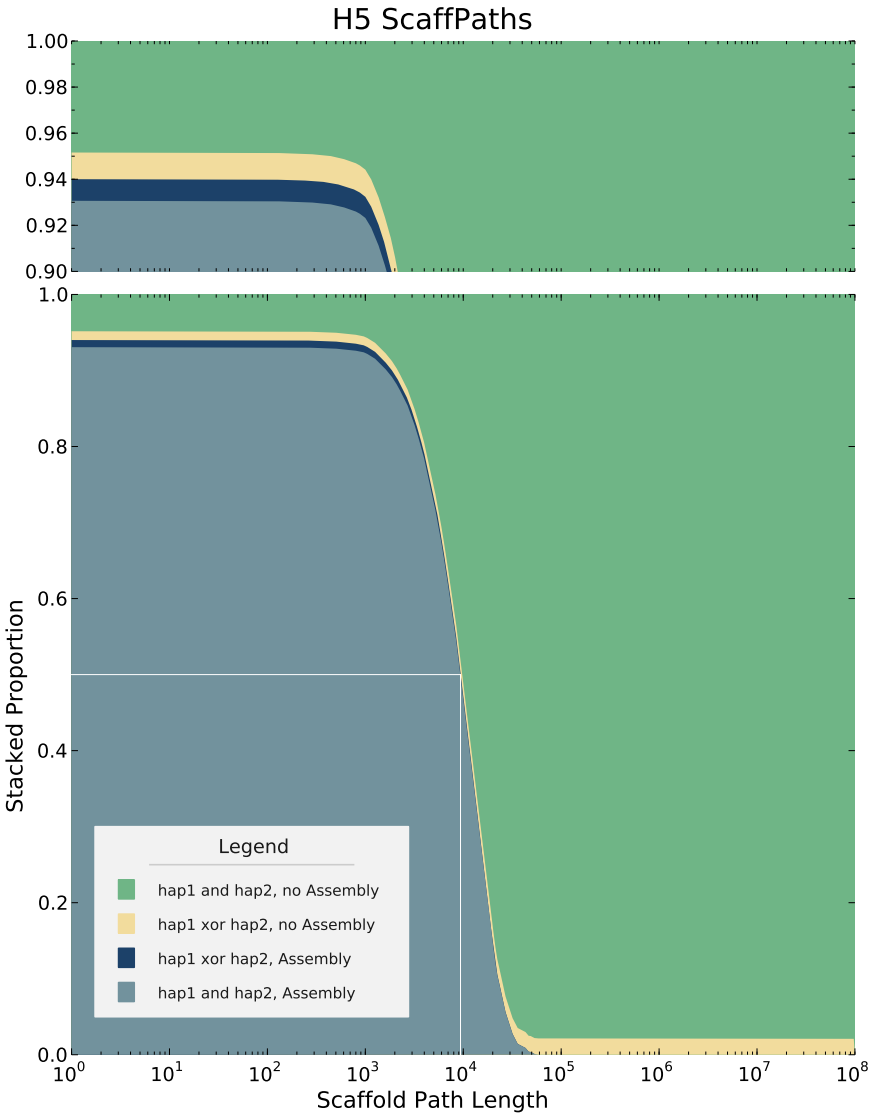


Figure 3.108: H5 scaffolds caption goes here.

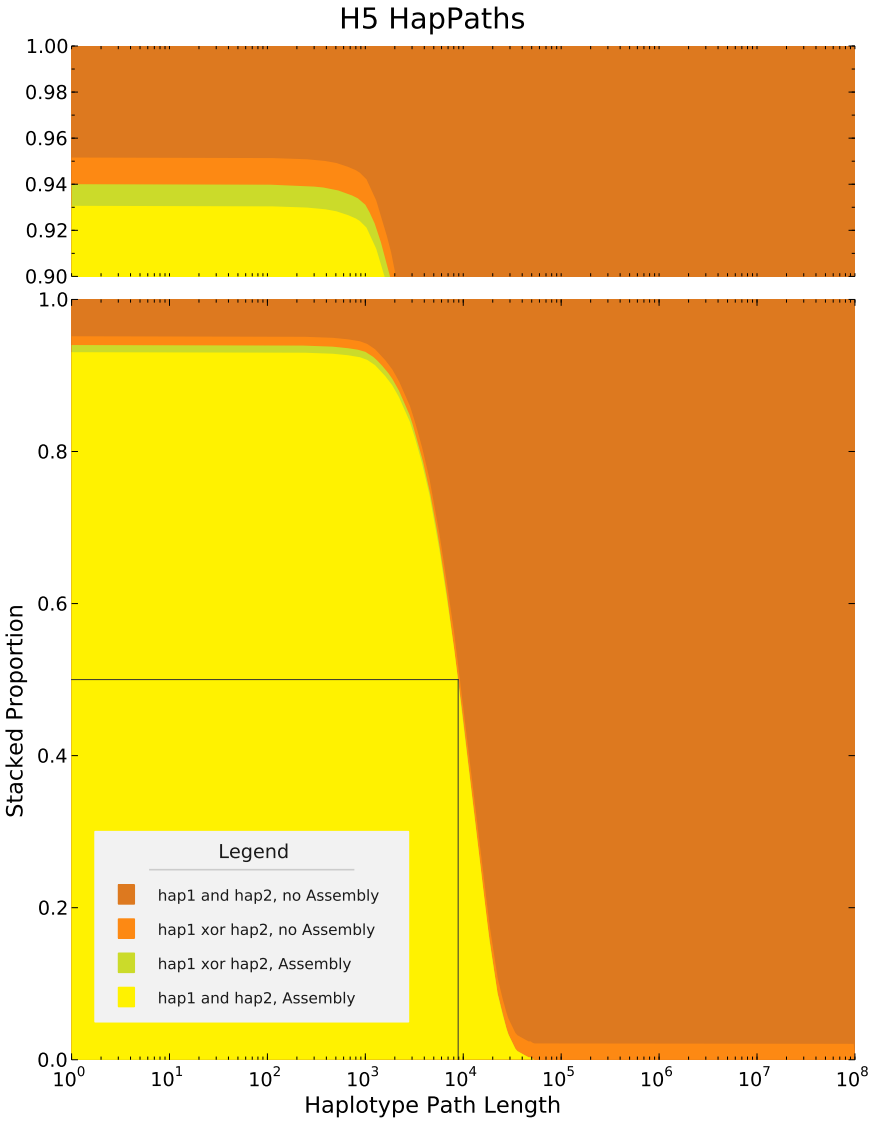


Figure 3.109: H5 hapPaths caption goes here.

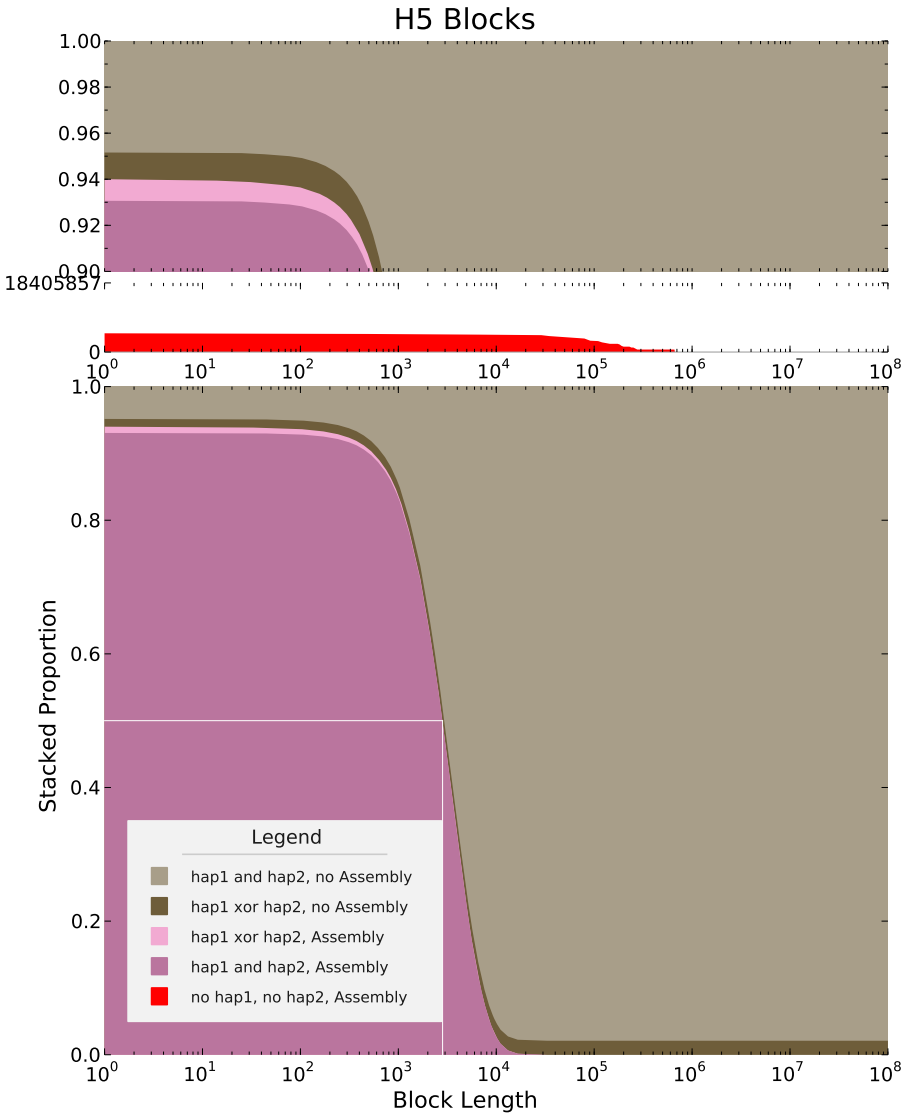


Figure 3.110: H5 blocks caption goes here.

3.2.9 I, Terrapins

Affiliation: CSHL (Cold Spring Harbor Laboratory), USA
Contact: Michael Schatz
Software: **Quake, Celera, Bambus2**
Number of entries: 2

ID	Total	Hap 1	Hap 2	Bac
I2	0.98467	0.98511	0.98424	0.99857
I1	0.87175	0.87213	0.87138	0.99691

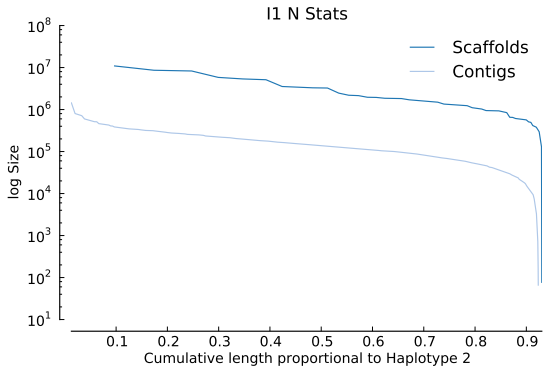
Assemblies:

I1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
A1	0.90867	0.90879	0.90854	0.99971
I1	0.87175	0.87213	0.87138	0.99691
N1	0.87107	0.87121	0.87094	0.00000

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	128	77	161.25	978	817,768.10	945,824.00	10,924,052	1,725,203.70	104,674,317
Contigs	1,798	64	1,007.25	20,276	57,785.43	79,792.25	1,442,666	92,952.55	103,898,208

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,274,593 – 109,998,671	95,859,650 – 96,440,001	191,713,066.0 – 192,858,436.0	1,292 – 3,779
Heterozygous	429,957 – 439,141	376,063 – 383,892	752,078.0 – 767,588.0	12 – 47
Indel	2,418,790 – 2,819,795	1,072,681 – 1,268,760	2,140,966.0 – 2,527,462.0	2,094 – 3,178

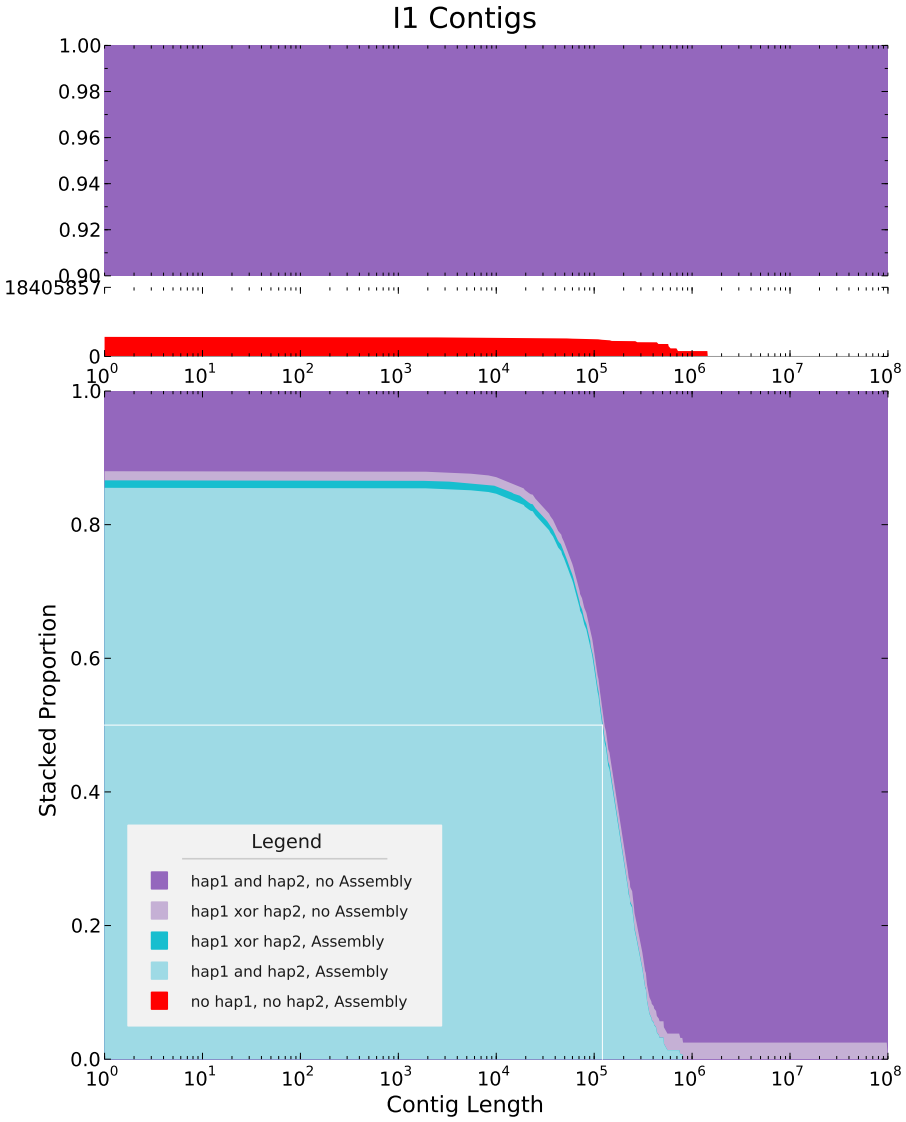


Figure 3.111: I1 contigs caption goes here.

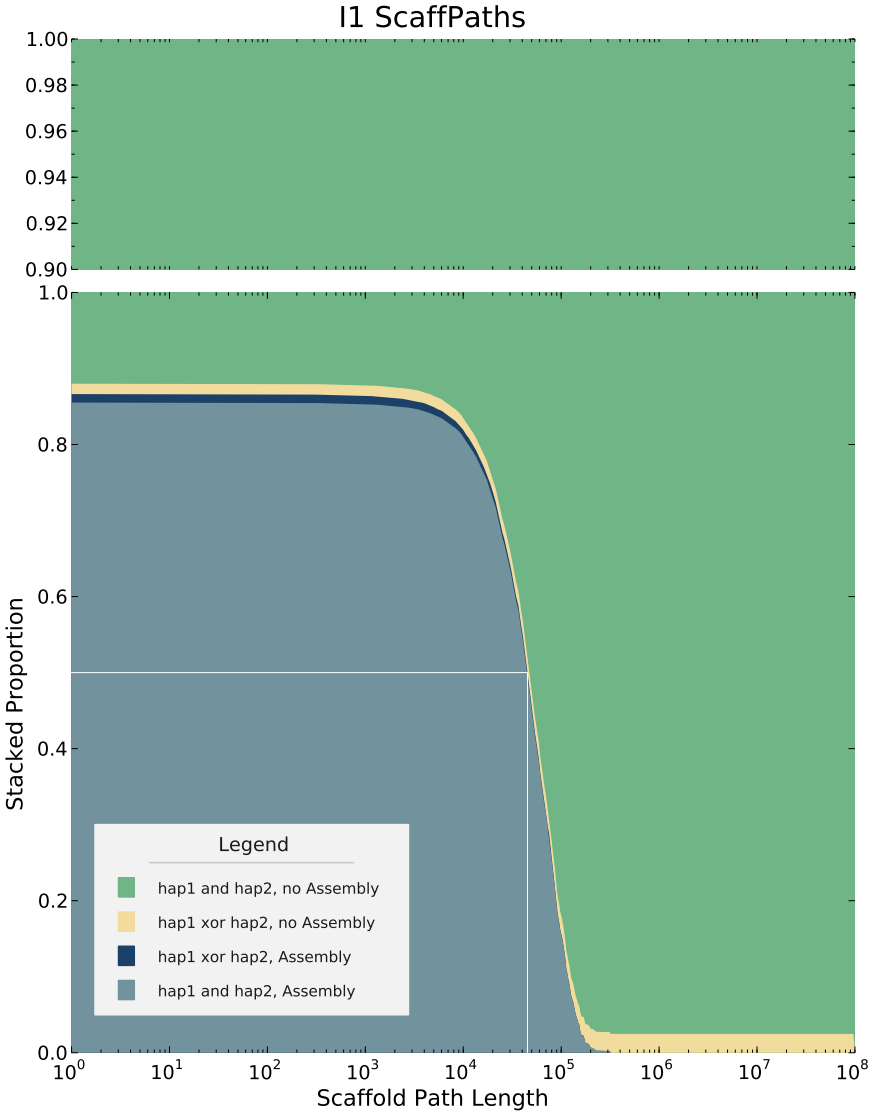


Figure 3.112: I1 scaffolds caption goes here.

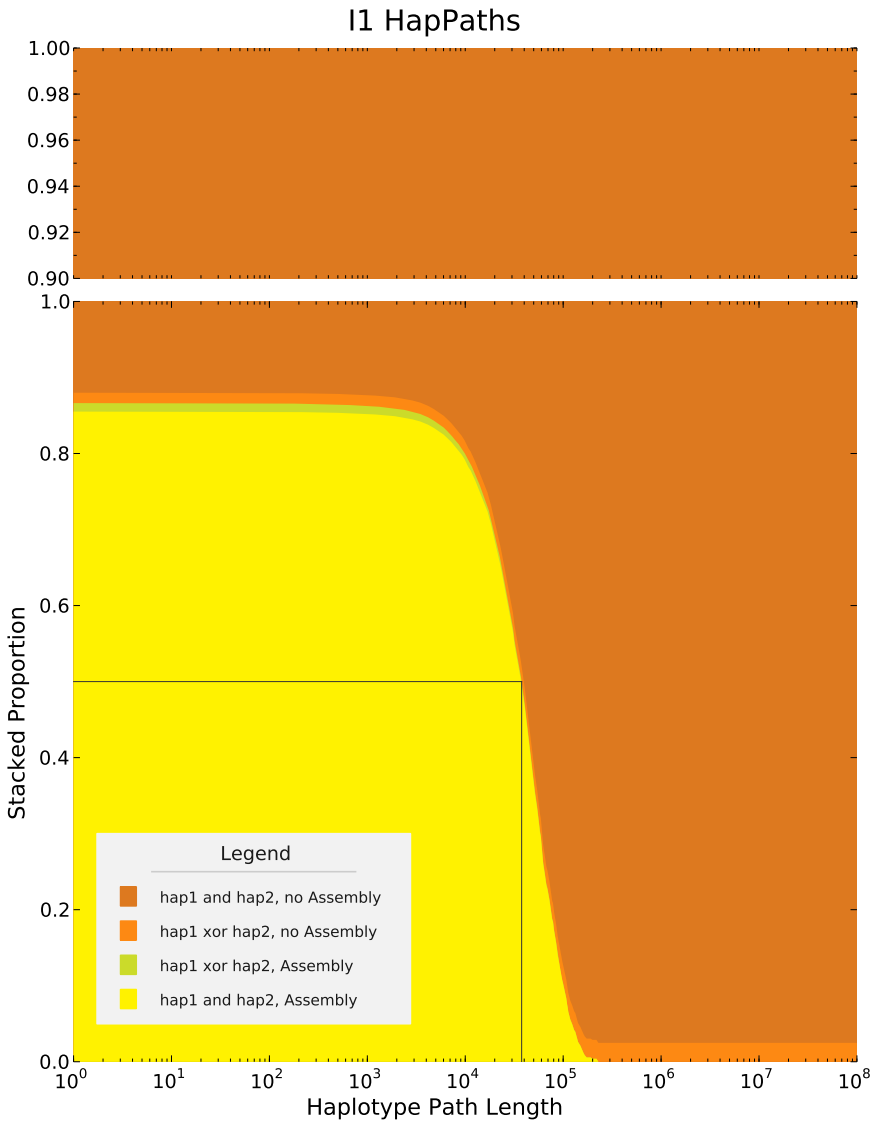


Figure 3.113: I1 hapPaths caption goes here.

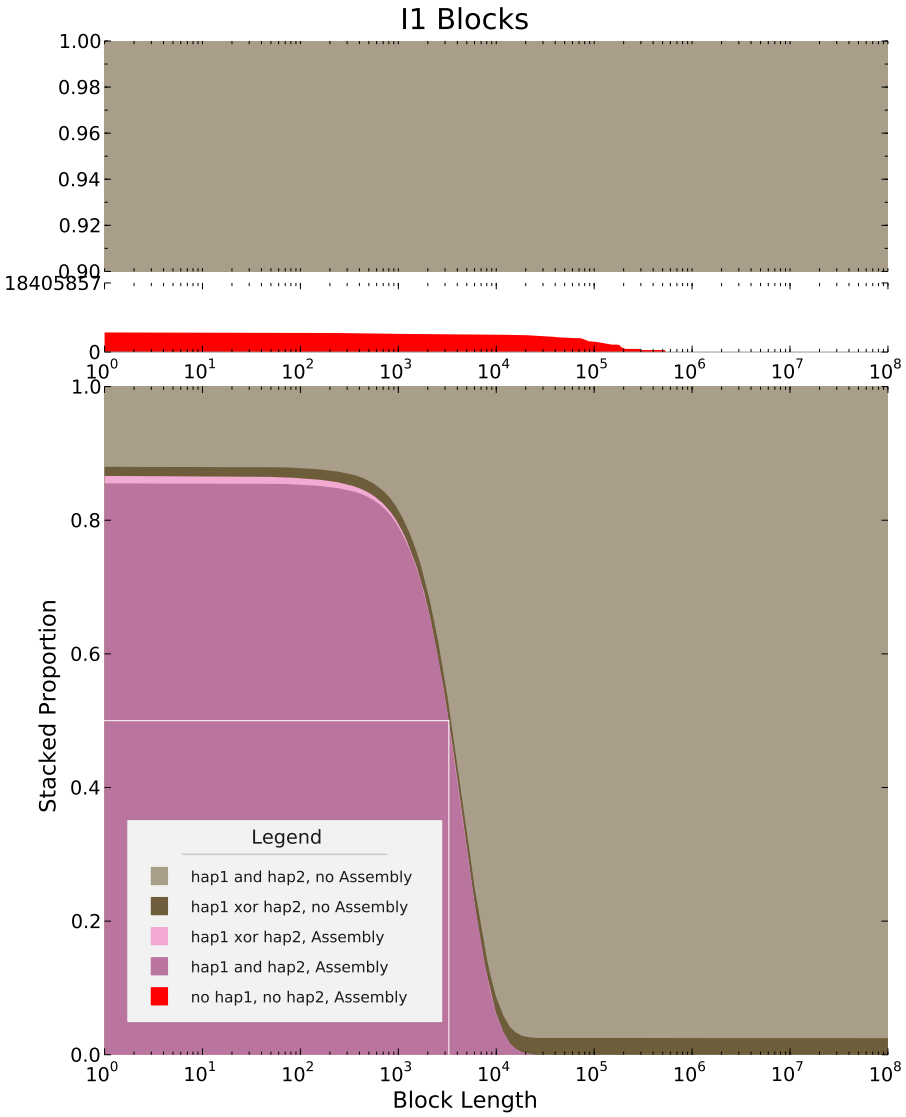


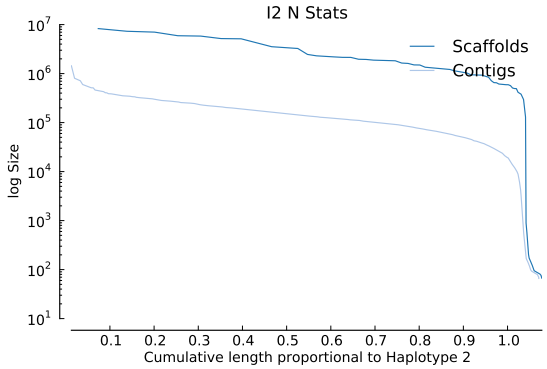
Figure 3.114: I1 blocks caption goes here.

I2

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
M1	0.98499	0.98522	0.98477	0.99967
I2	0.98467	0.98511	0.98424	0.99857
Q1	0.98325	0.98337	0.98311	0.68864

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	35,661	64	83.00	89	3,400.34	115.00	8,283,751	110,072.92	121,259,411
Contigs	37,571	64	83.00	90	3,207.30	126.00	1,442,666	24,761.54	120,501,333

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,081,096 – 108,814,321	107,221,079 – 107,873,260	214,436,520.0 – 215,728,370.0	1,471 – 4,204
Heterozygous	423,181 – 432,199	419,081 – 426,660	838,118.0 – 853,016.0	15 – 127
Indel	2,701,965 – 3,143,423	1,249,564 – 1,577,816	2,494,016.0 – 3,144,390.0	2,436 – 3,696

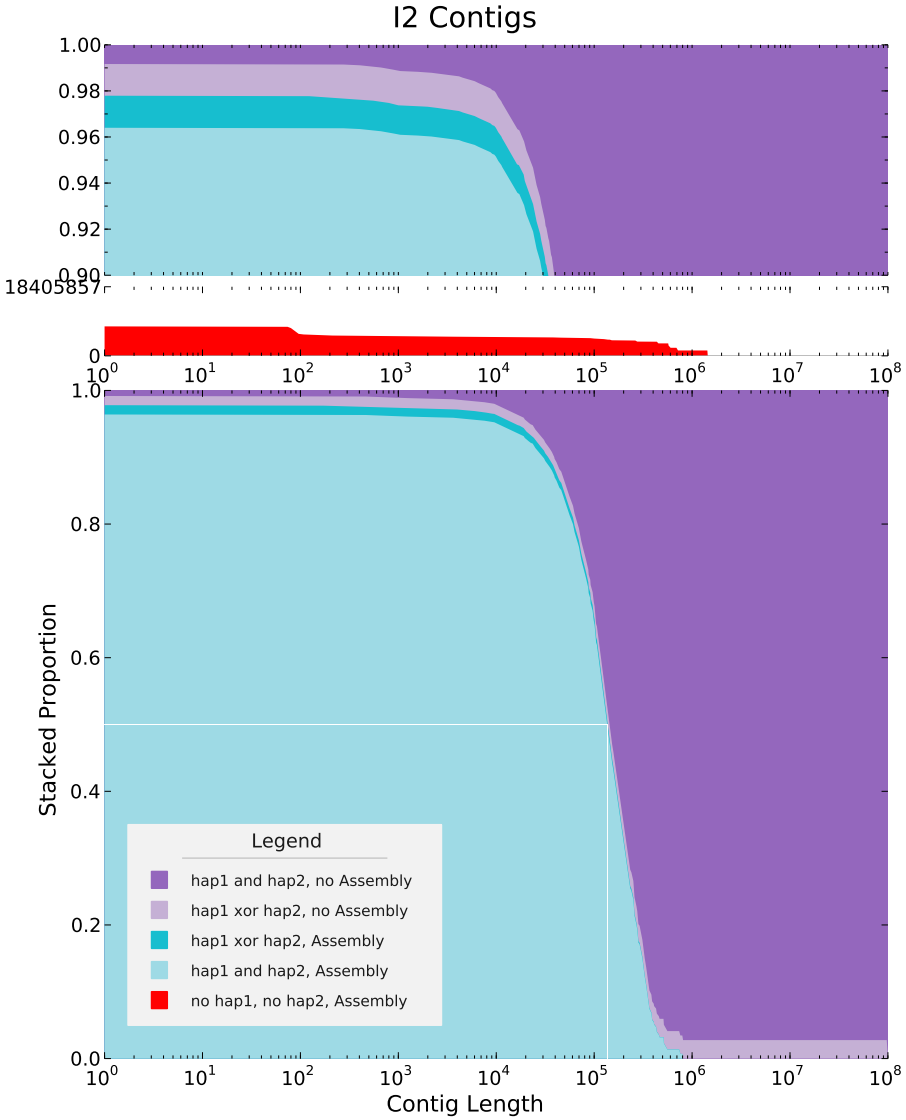


Figure 3.115: I2 contigs caption goes here.

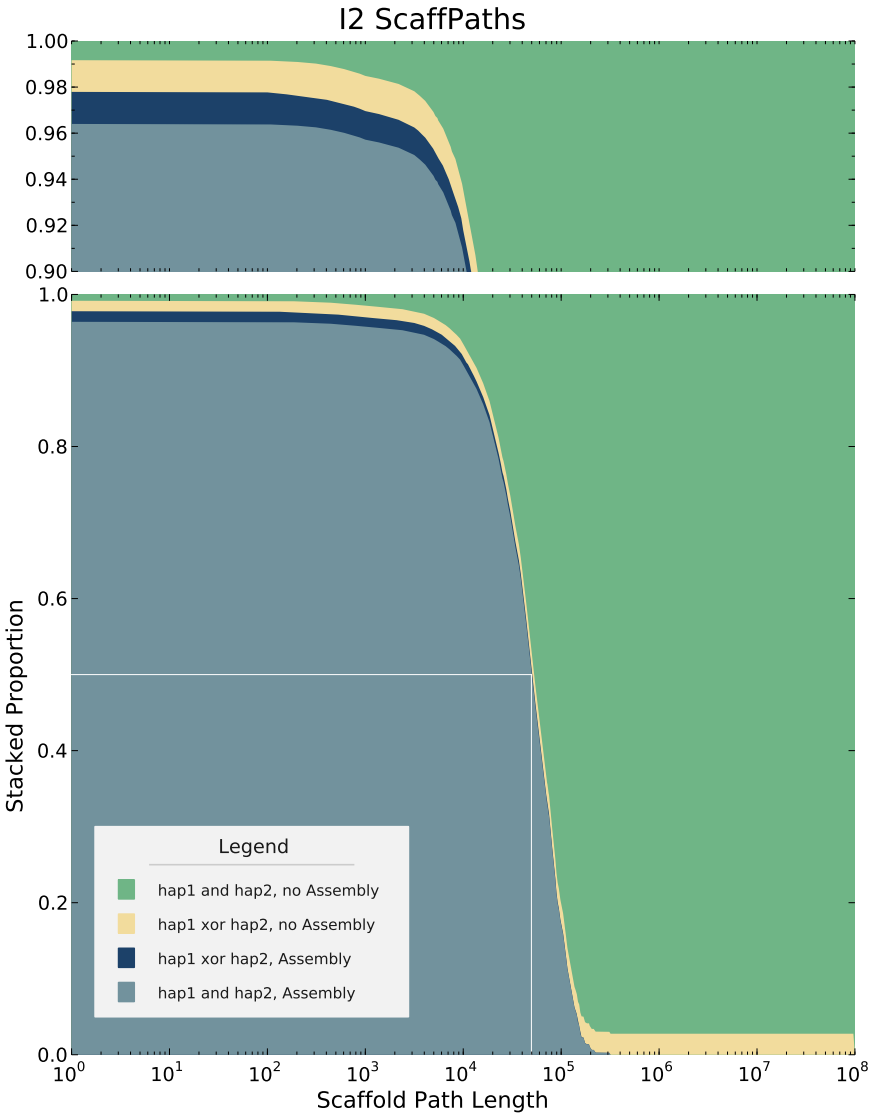


Figure 3.116: I2 scaffolds caption goes here.

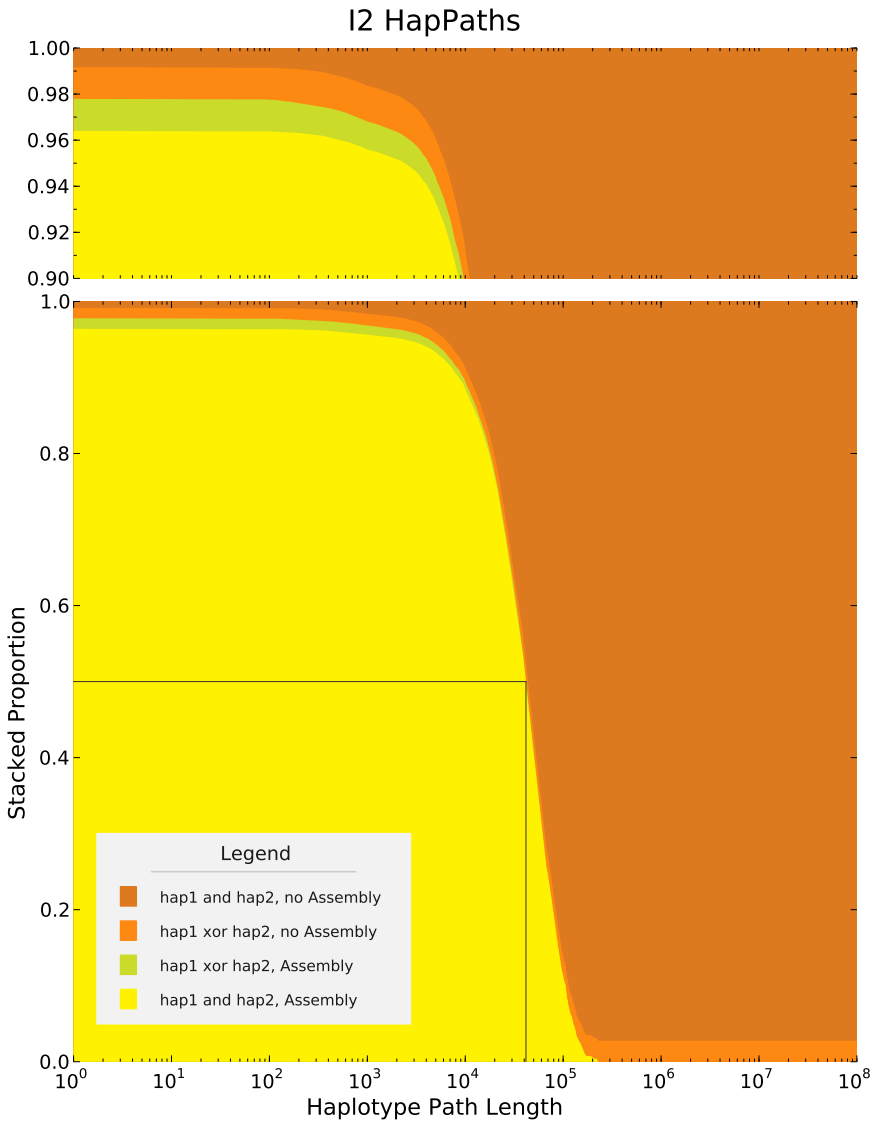


Figure 3.117: I2 hapPaths caption goes here.

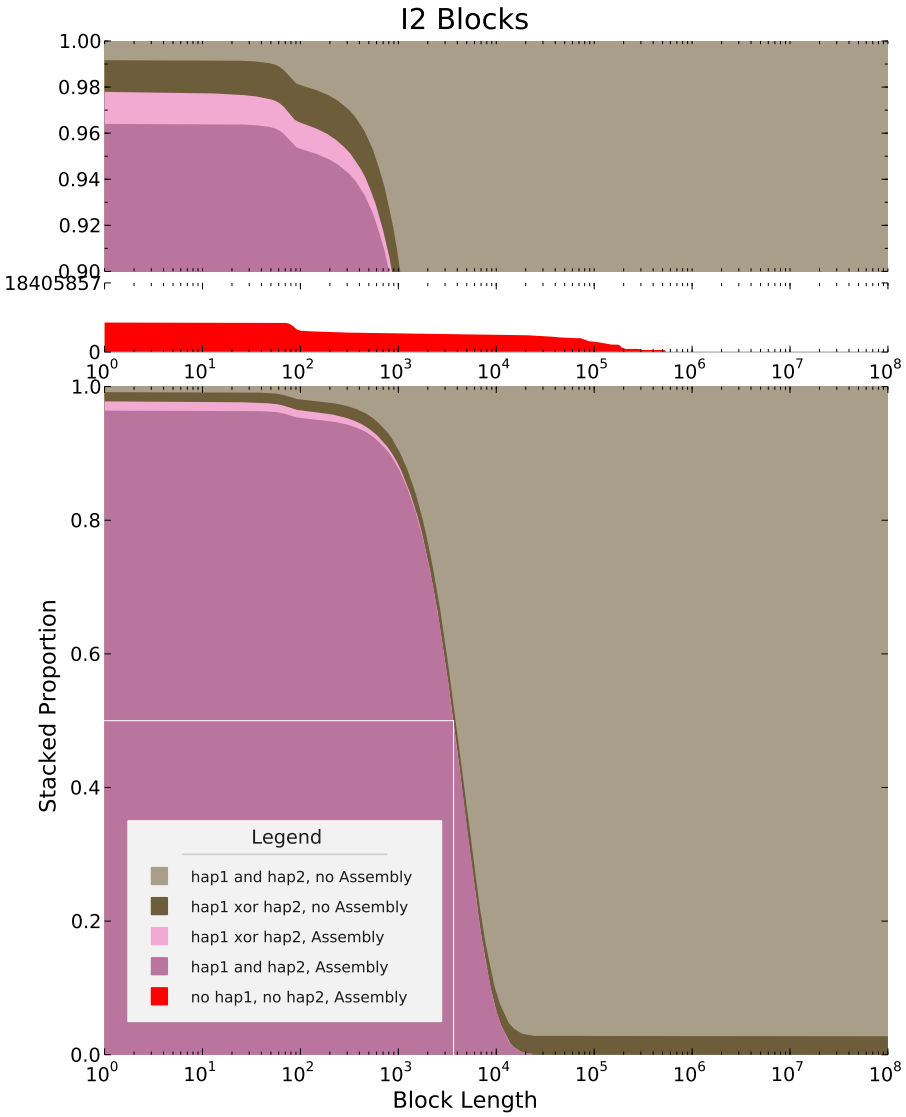


Figure 3.118: I2 blocks caption goes here.

3.2.10 J, Xiaoqiu Huang

Affiliation: Department of Computer Science, Iowa State University
Contact: Xiaoqiu Huang
Software: **PCAP**
Number of entries: 1

ID	Total	Hap 1	Hap 2	Bac
J1	0.94234	0.94249	0.94219	0.99517

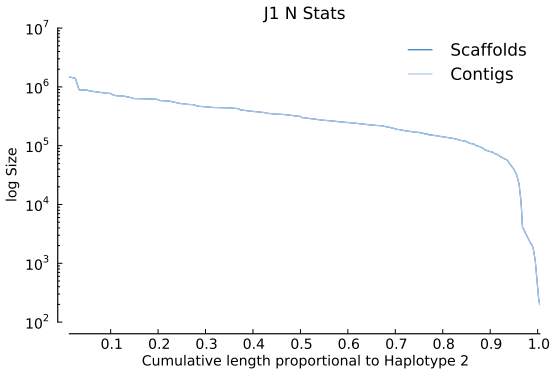
Assemblies:

J1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
H5	0.94518	0.94532	0.94503	0.99789
J1	0.94234	0.94249	0.94219	0.99517
H1	0.93573	0.93582	0.93567	0.99709

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	4,791	200	245.00	556	23,594.91	2,233.00	1,474,238	90,842.09	113,043,196
Contigs	4,791	200	245.00	556	23,594.91	2,233.00	1,474,238	90,842.09	113,043,196

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,307,624 – 110,037,419	103,578,656 – 104,192,856	207,132,892.0 – 208,305,362.0	3,010 – 8,540
Heterozygous	432,048 – 439,528	407,486 – 414,252	814,786.0 – 828,080.0	14 – 31
Indel	1,771,527 – 2,157,718	757,261 – 983,039	1,511,650.0 – 1,956,348.0	1,148 – 2,255

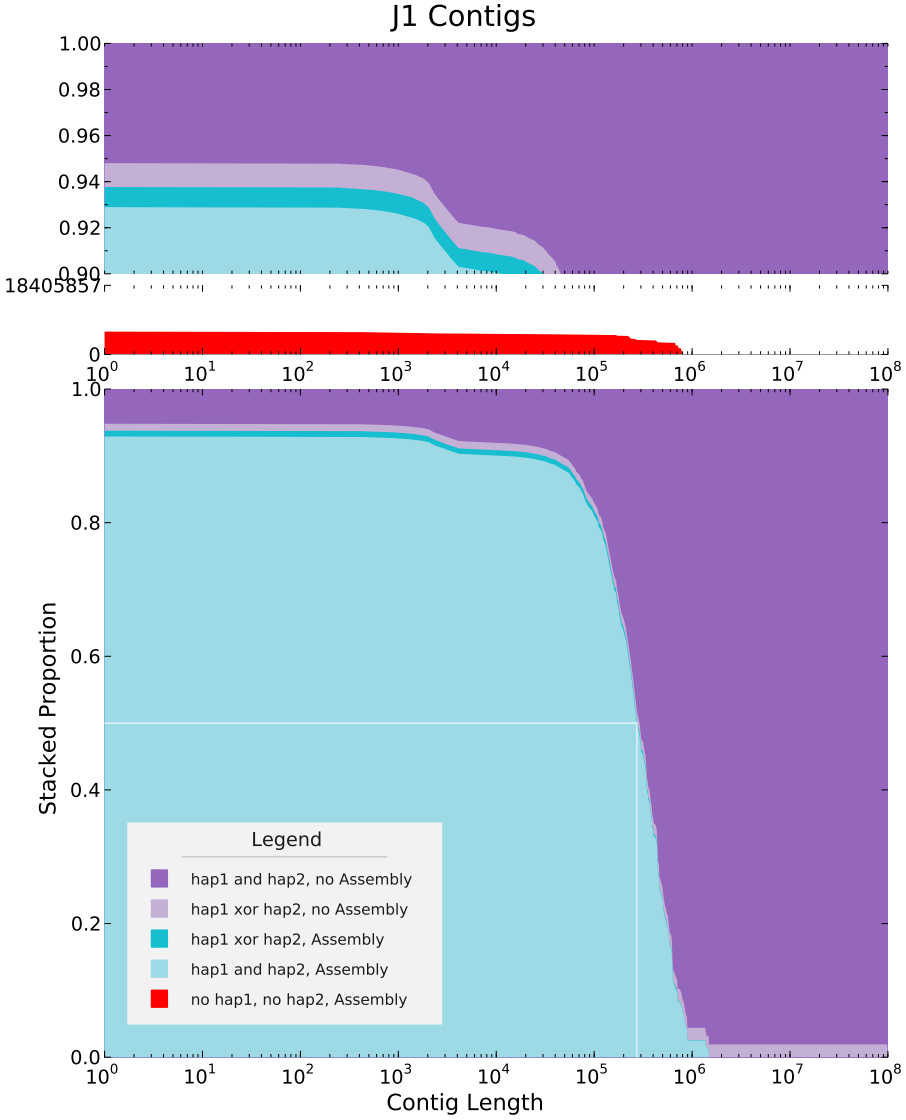


Figure 3.119: J1 contigs caption goes here.

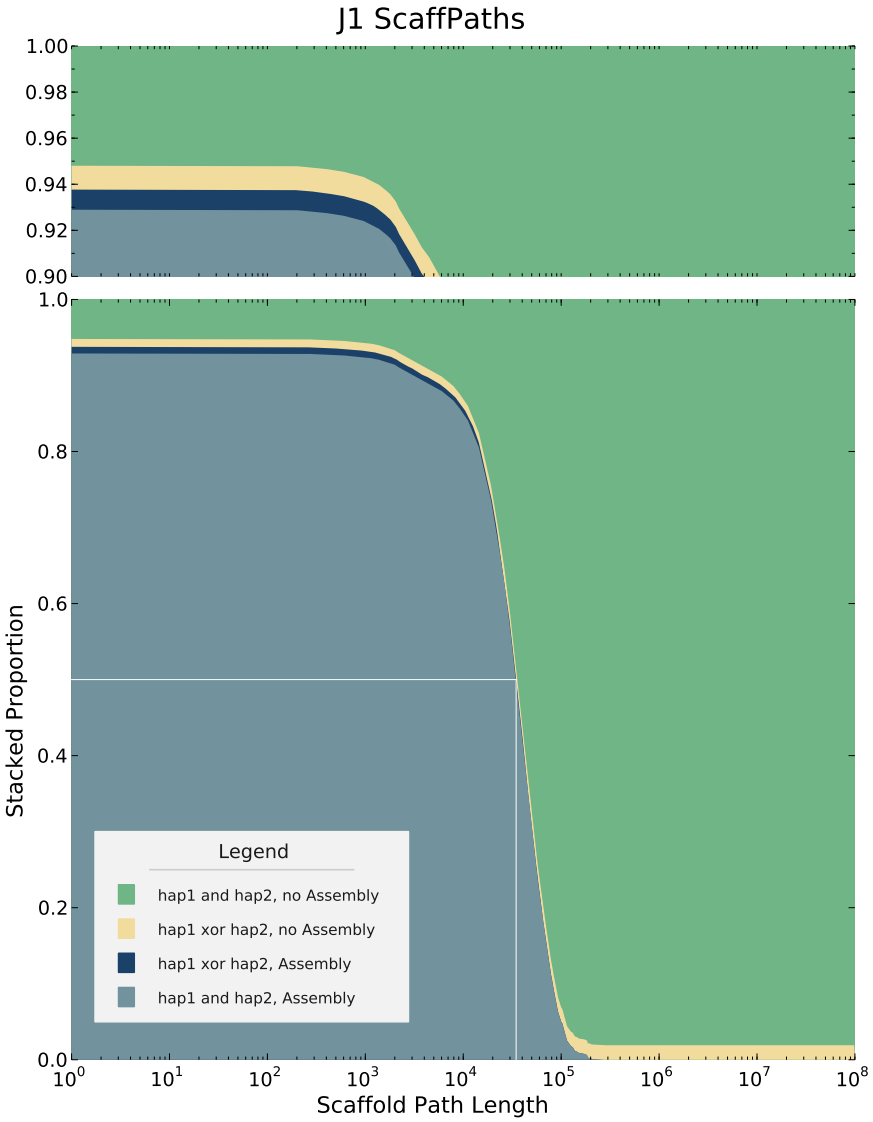


Figure 3.120: J1 scaffolds caption goes here.

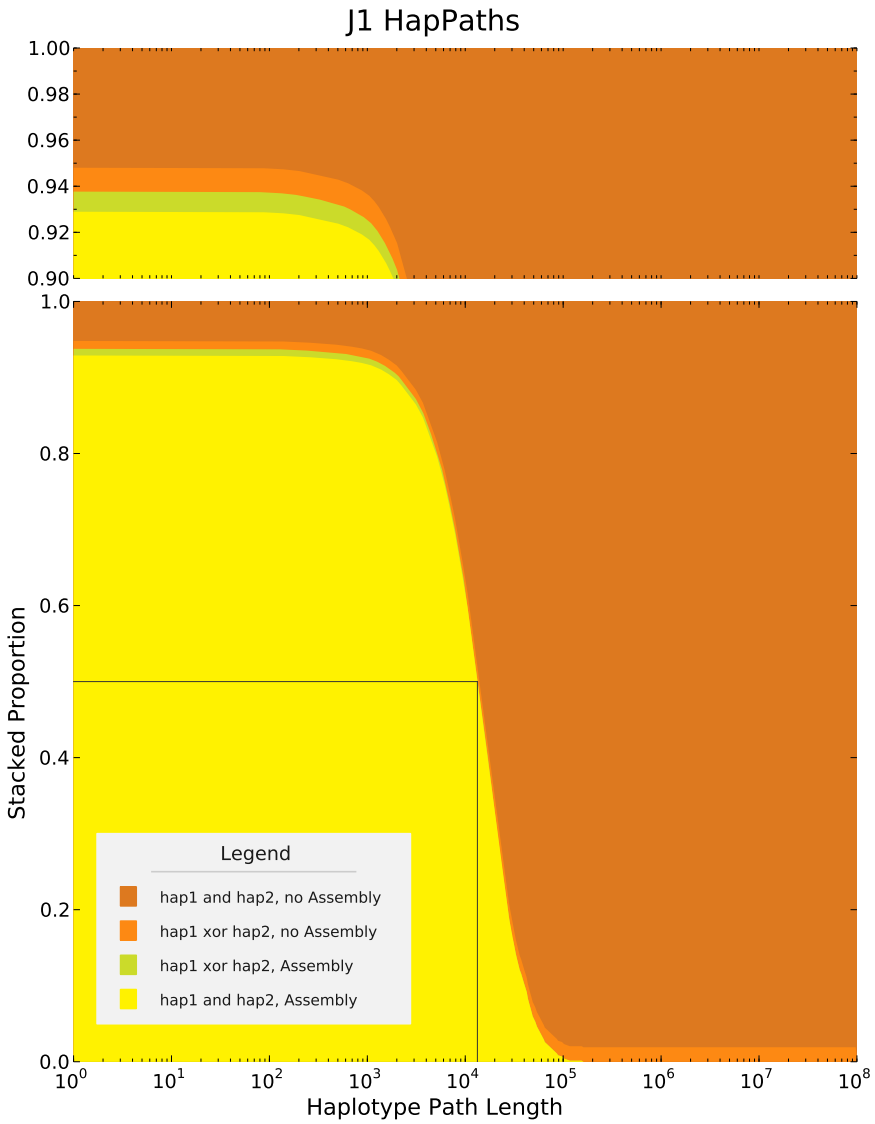


Figure 3.121: J1 hapPaths caption goes here.

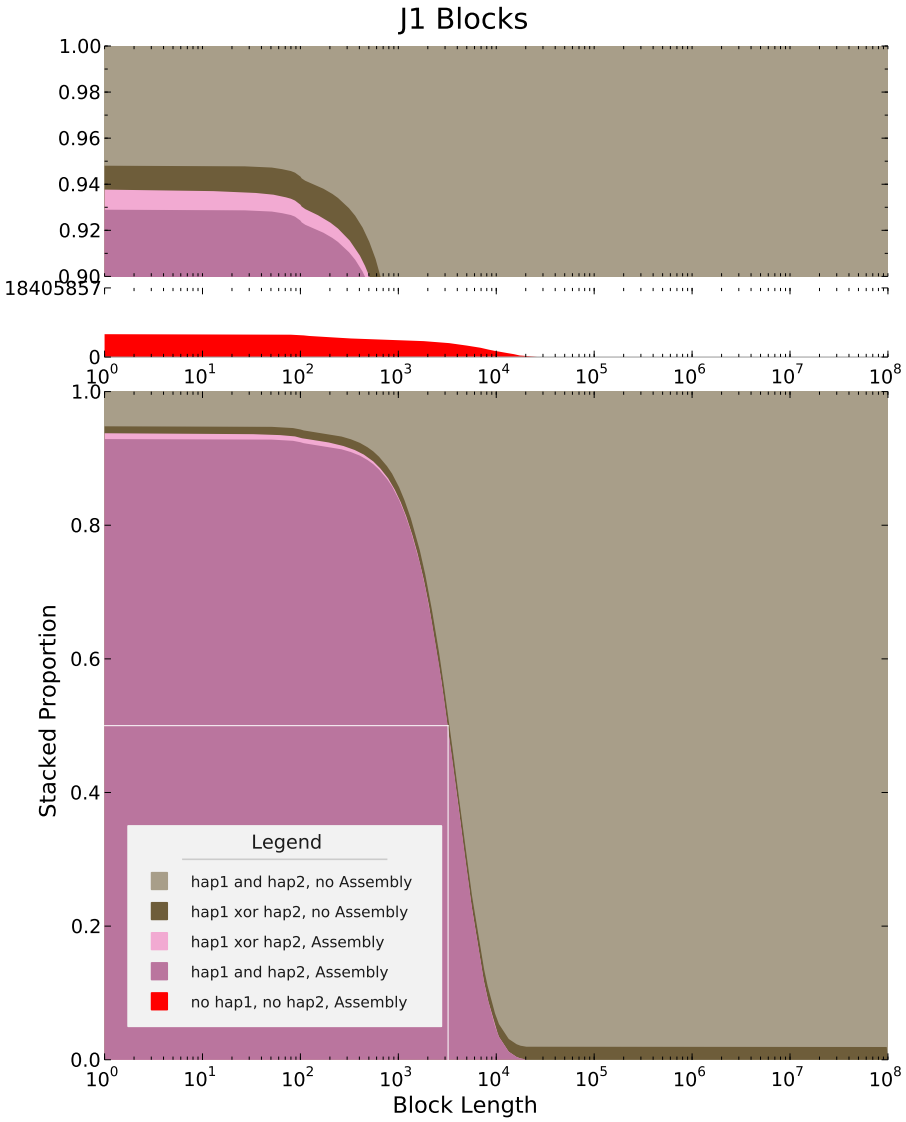


Figure 3.122: J1 blocks caption goes here.

3.2.11 K, Super Dawgs

Affiliation: Computational Systems Biology Laboratory, University of Georgia, USA
Contact: Wen-Chi Chou
Software: **Seqclean, SOAPdenovo**
Number of entries: 3

ID	Total	Hap 1	Hap 2	Bac
K1	0.98306	0.98309	0.98302	0.11258
K2	0.98288	0.98287	0.98288	0.04786
K3	0.98038	0.98042	0.98034	0.00017

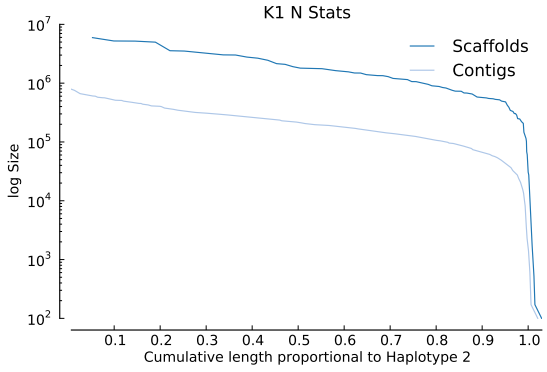
Assemblies:

K1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
Q1	0.98325	0.98337	0.98311	0.68864
K1	0.98306	0.98309	0.98302	0.11258
D4	0.98288	0.98303	0.98274	0.99618

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	14,856	100	111.00	128	7,799.74	153.00	5,949,022	135,543.40	115,872,926
Contigs	15,689	100	112.00	131	7,325.25	158.00	788,375	42,135.61	114,925,837

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	107,089,490 – 107,970,124	106,231,521 – 106,945,316	212,452,692.0 – 213,870,430.0	5,150 – 9,996
Heterozygous	419,641 – 431,126	415,464 – 425,503	830,896.0 – 850,928.0	15 – 38
Indel	2,772,237 – 3,164,218	1,231,746 – 1,434,638	2,459,500.0 – 2,862,004.0	1,995 – 3,626

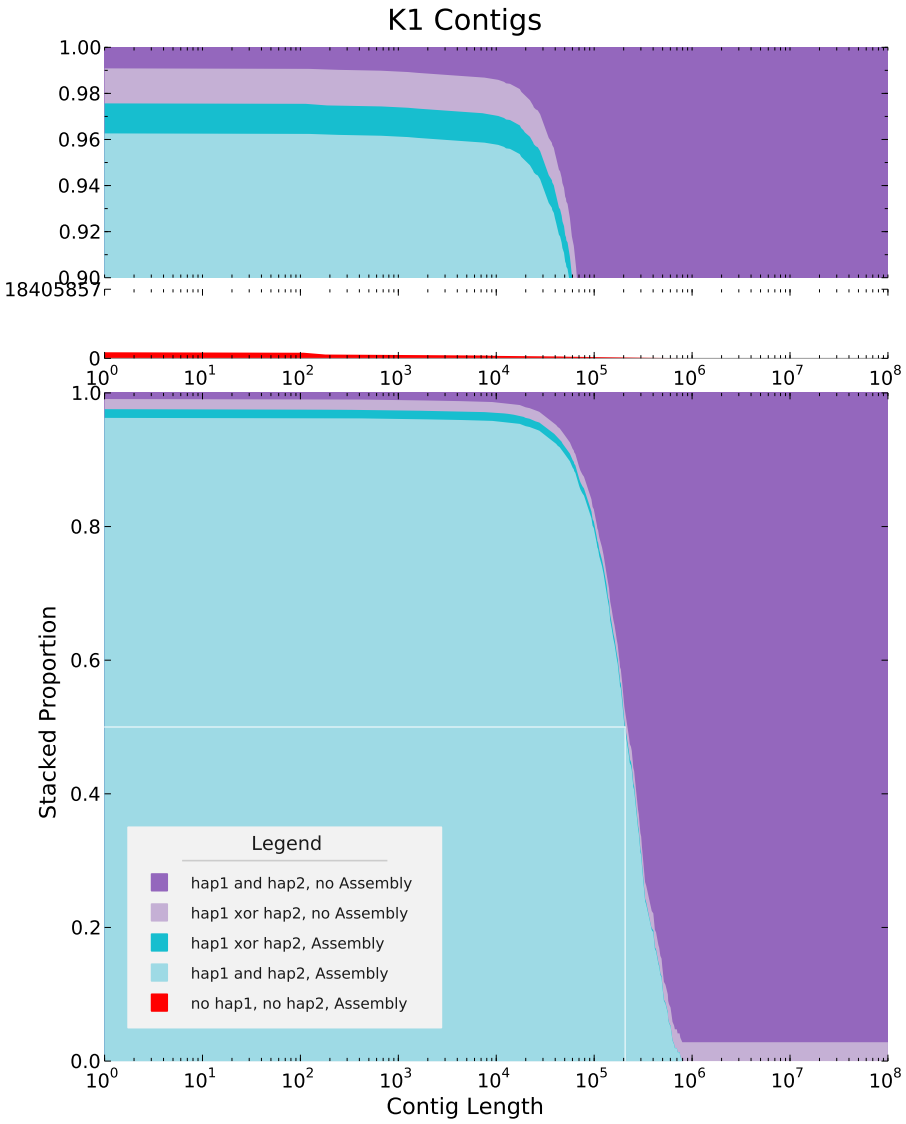


Figure 3.123: K1 contigs caption goes here.

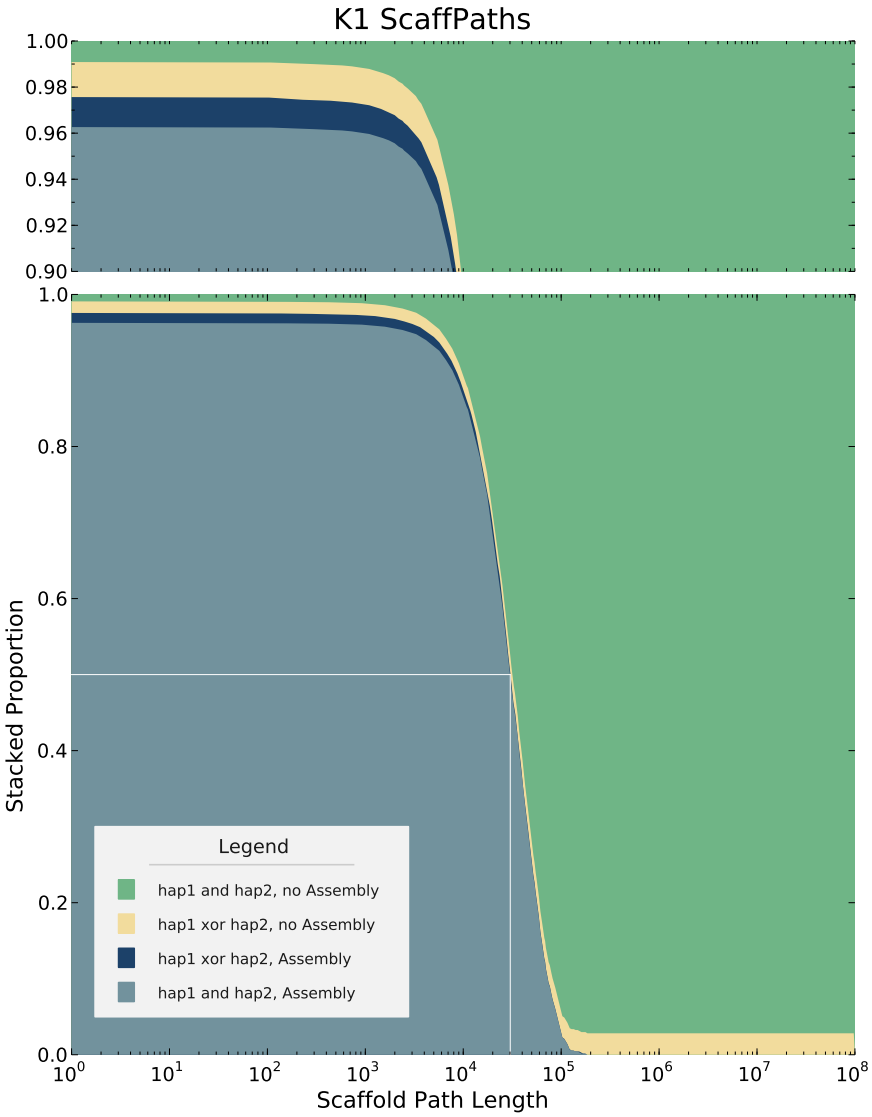


Figure 3.124: K1 scaffolds caption goes here.

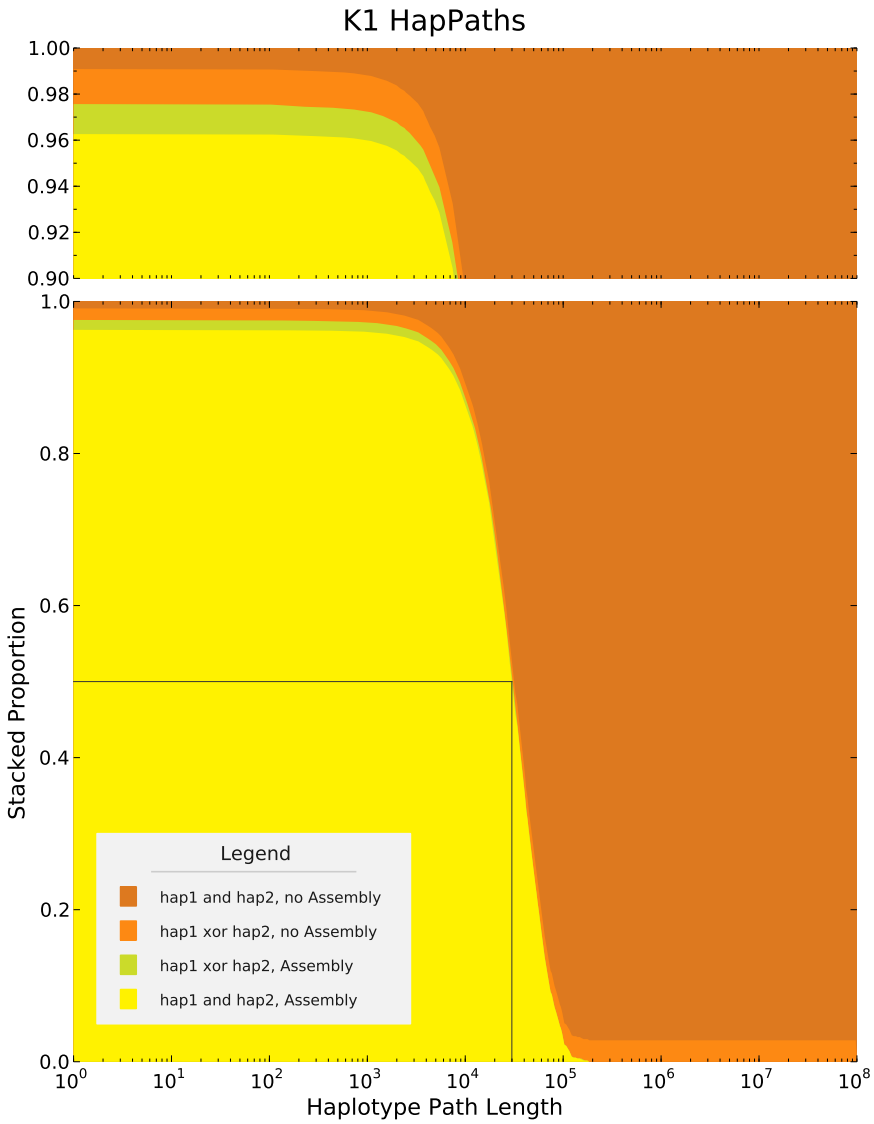


Figure 3.125: K1 hapPaths caption goes here.

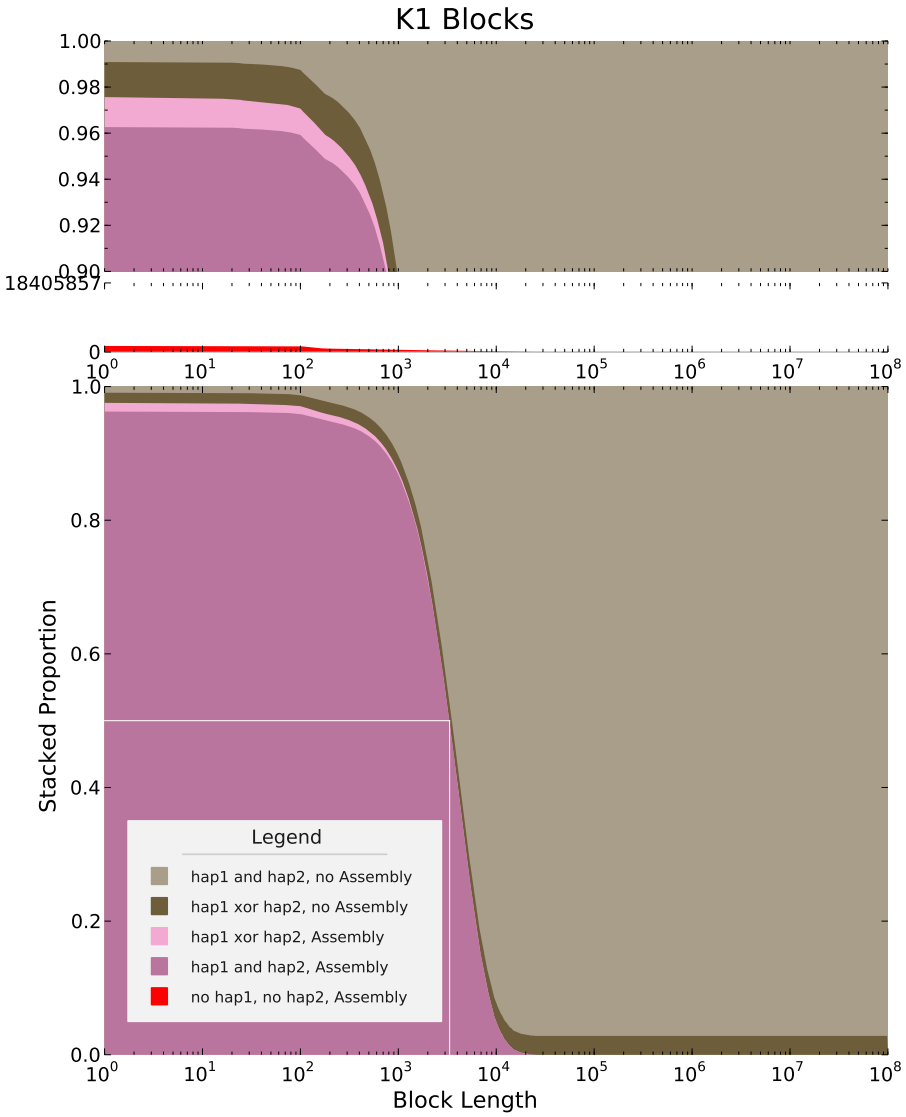


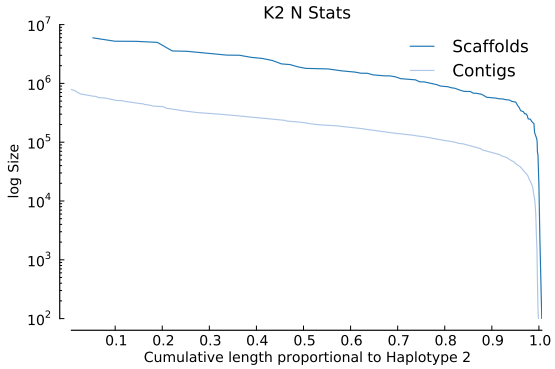
Figure 3.126: K1 blocks caption goes here.

K2

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
D4	0.98288	0.98303	0.98274	0.99618
K2	0.98288	0.98287	0.98288	0.04786
K3	0.98038	0.98042	0.98034	0.00017

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	2,043	100	112.00	132	55,420.90	173.00	5,949,022	361,960.27	113,224,906
Contigs	2,796	100	117.00	159	40,217.11	27,335.75	788,375	92,986.88	112,447,027

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,550,808 – 109,349,358	107,670,972 – 108,302,410	215,331,006.0 – 216,584,434.0	5,444 – 10,088
Heterozygous	426,211 – 436,207	421,963 – 430,494	843,894.0 – 860,908.0	15 – 39
Indel	2,793,304 – 3,185,337	1,266,635 – 1,459,716	2,529,222.0 – 2,912,094.0	2,023 – 3,659

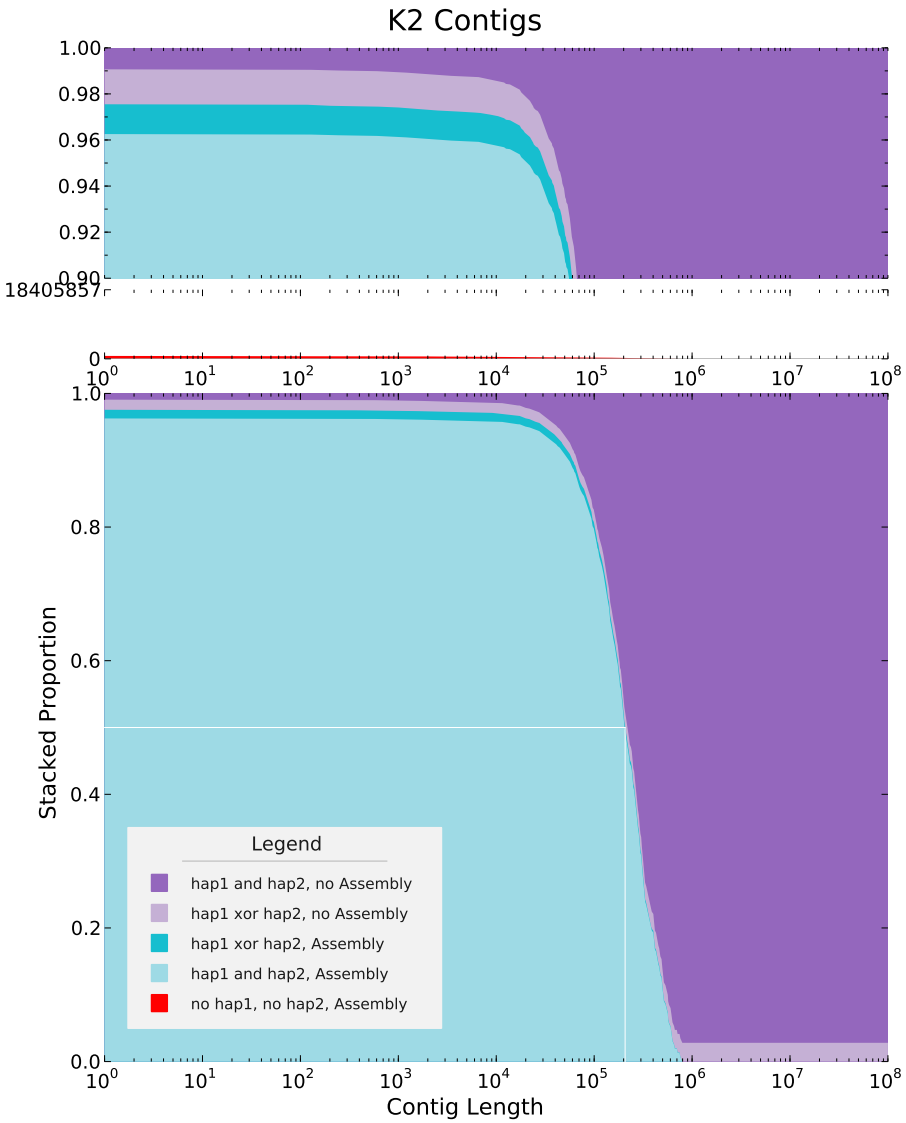


Figure 3.127: K2 contigs caption goes here.

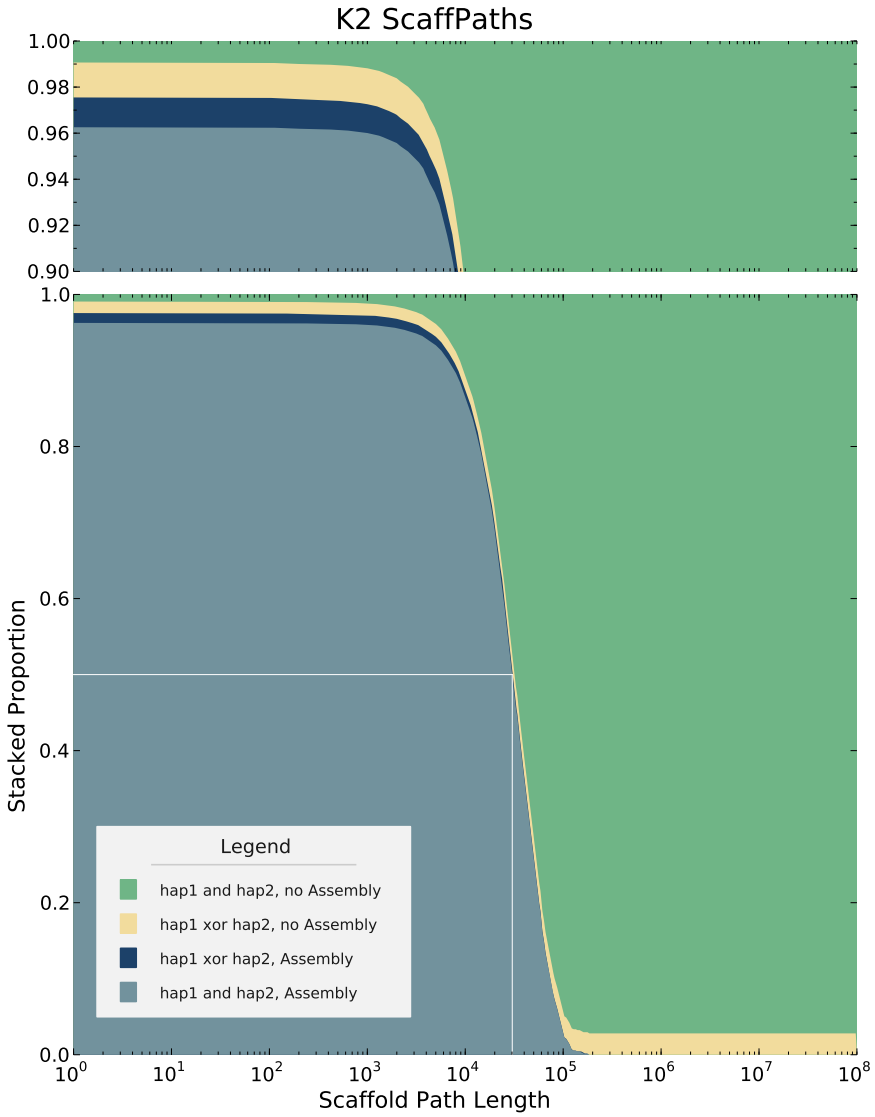


Figure 3.128: K2 scaffolds caption goes here.

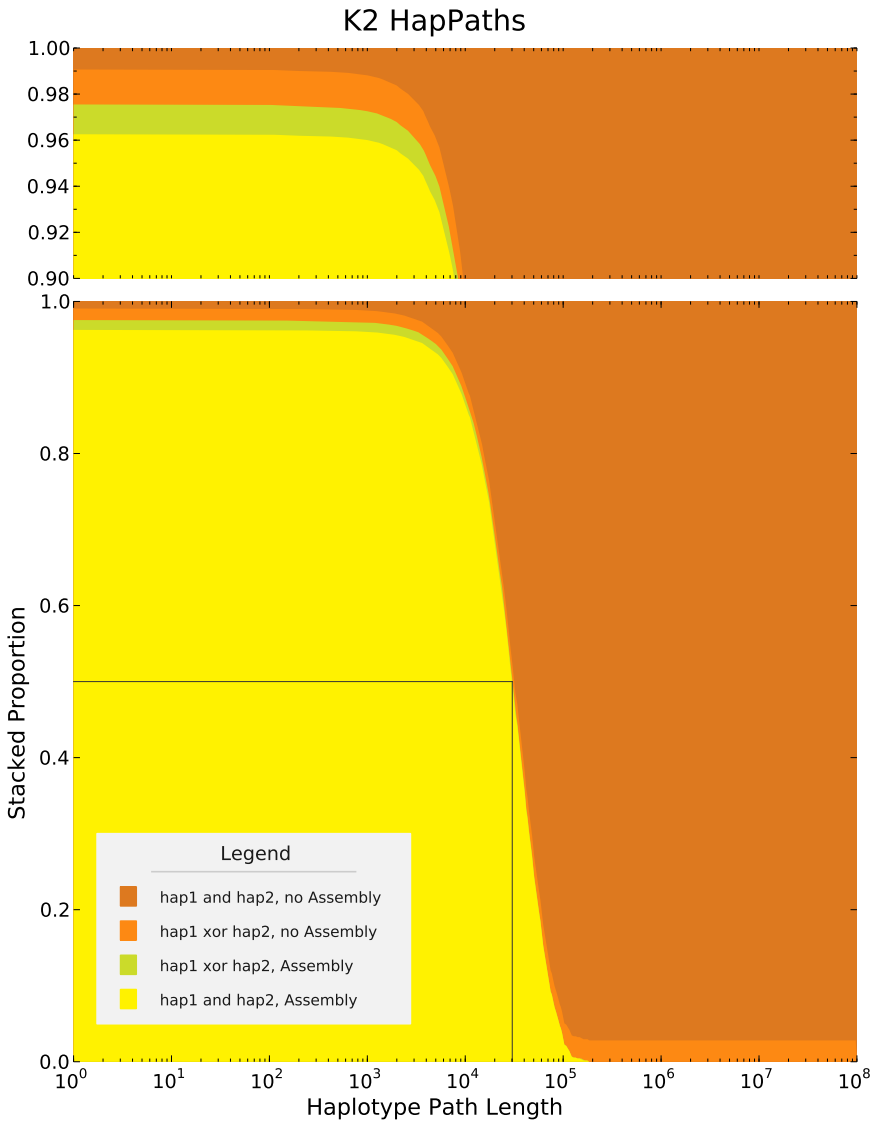


Figure 3.129: K2 hapPaths caption goes here.

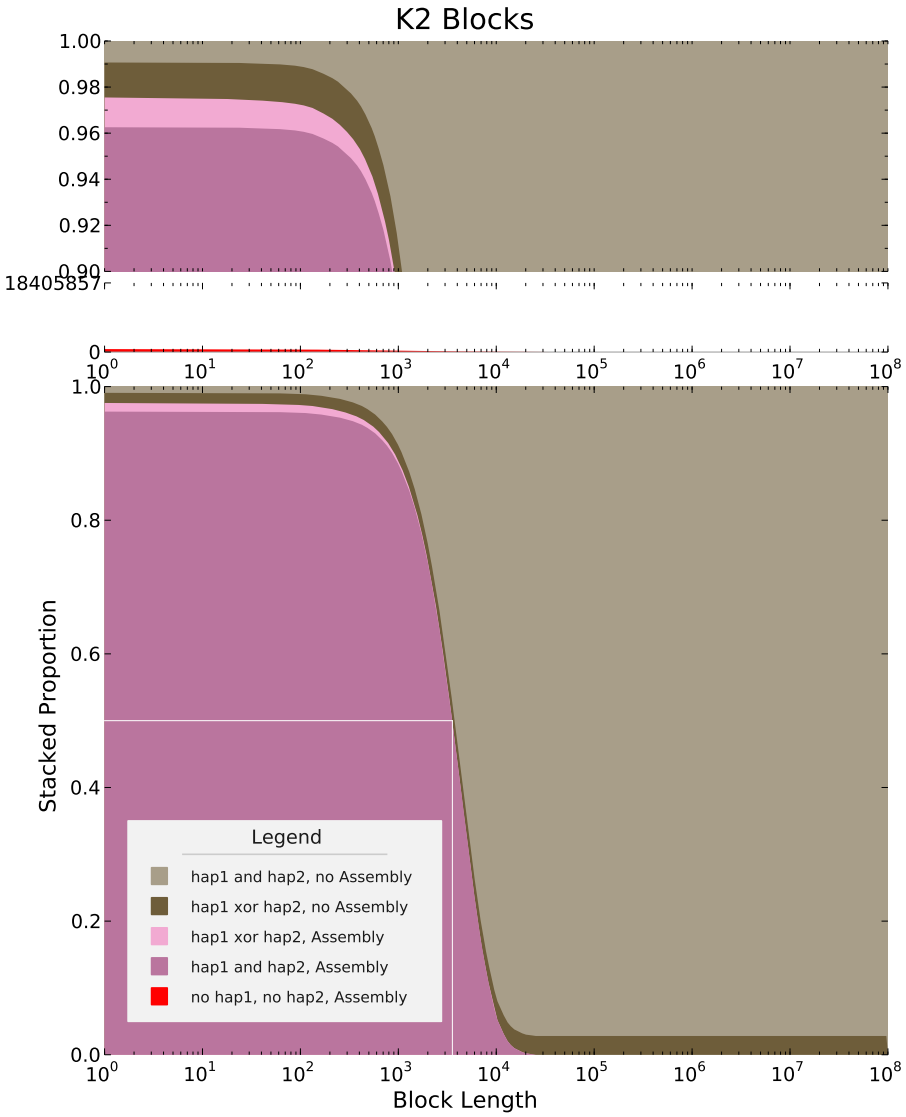


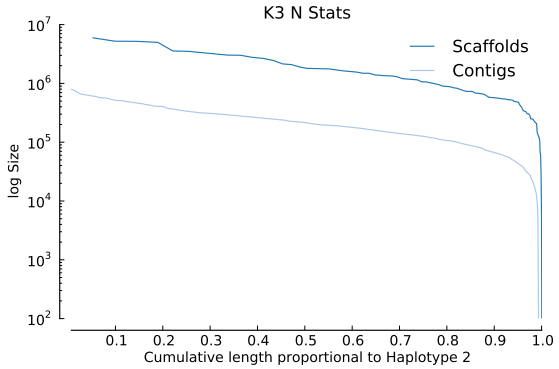
Figure 3.130: K2 blocks caption goes here.

K3

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
K2	0.98288	0.98287	0.98288	0.04786
K3	0.98038	0.98042	0.98034	0.00017
D2	0.97705	0.97715	0.97696	0.99086

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	173	100	769.00	131,182	650,799.61	843,108.00	5,949,022	1,079,762.10	112,588,332
Contigs	926	100	27,856.25	81,720	120,745.63	172,921.25	788,375	128,128.02	111,810,453

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,611,132 – 109,385,076	107,458,620 – 108,075,269	214,906,182.0 – 216,129,838.0	5,503 – 10,245
Heterozygous	426,783 – 436,632	421,408 – 429,919	842,784.0 – 859,760.0	15 – 38
Indel	2,794,782 – 3,185,607	1,263,667 – 1,440,996	2,523,274.0 – 2,874,628.0	2,029 – 3,672

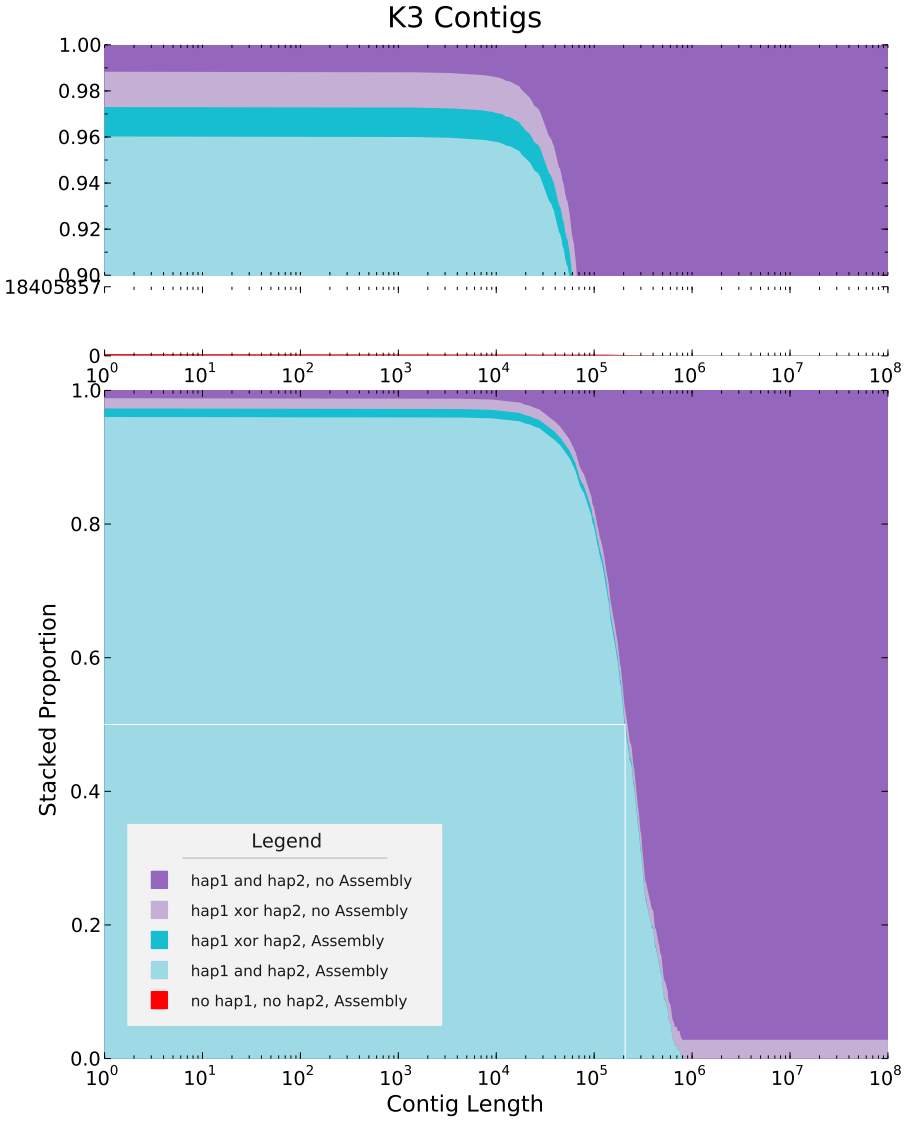


Figure 3.131: K3 contigs caption goes here.

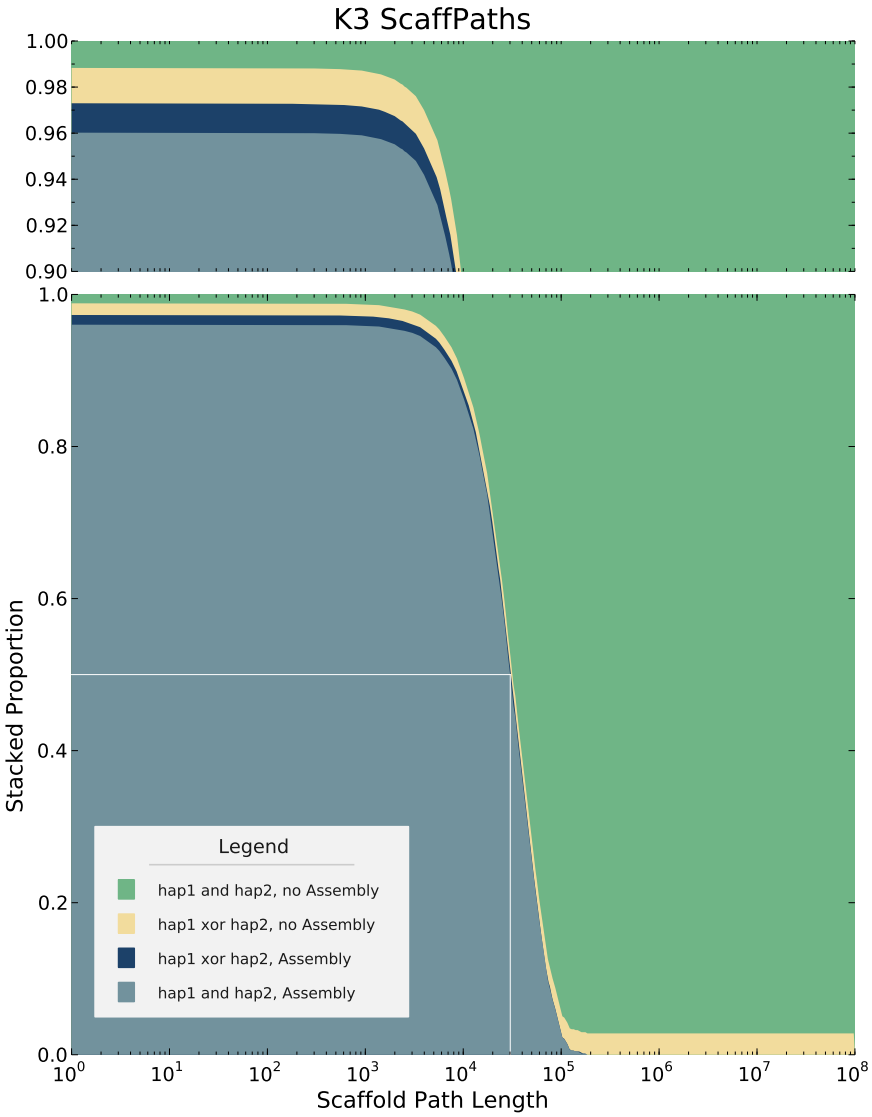


Figure 3.132: K3 scaffolds caption goes here.

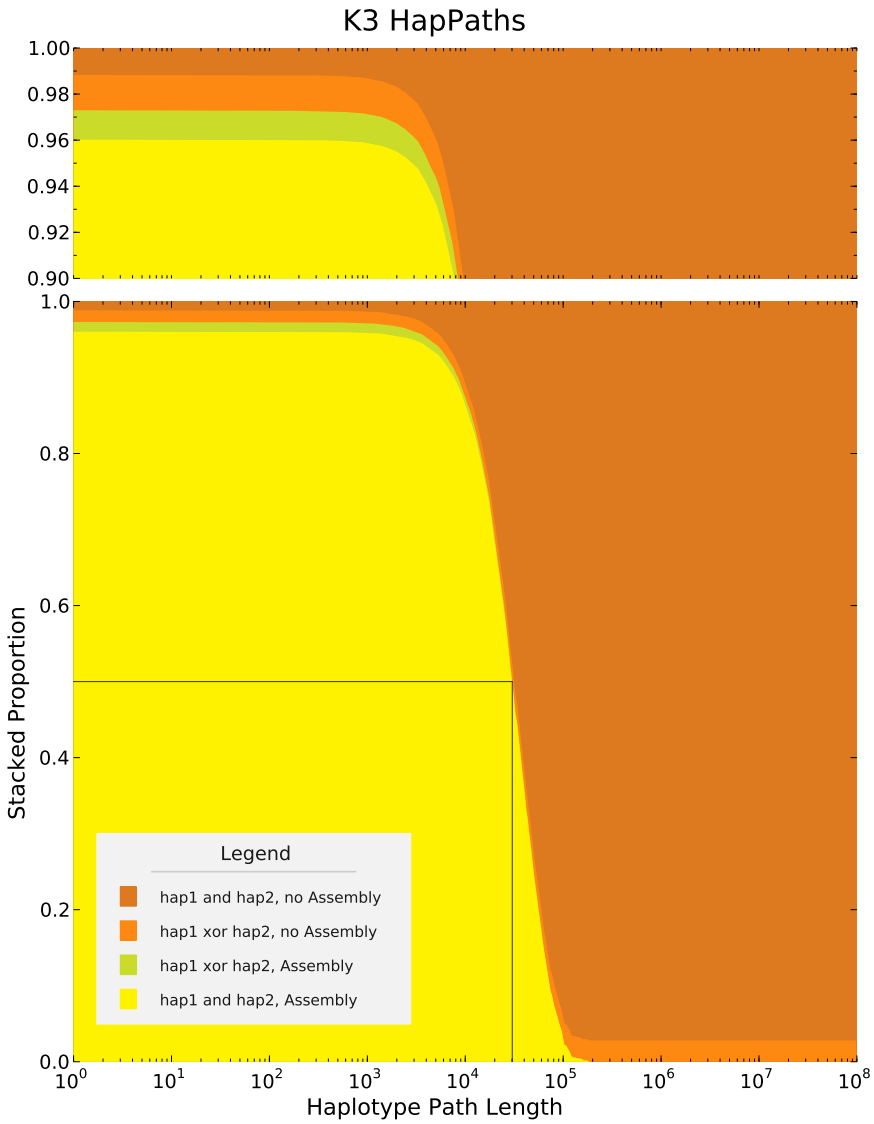


Figure 3.133: K3 hapPaths caption goes here.

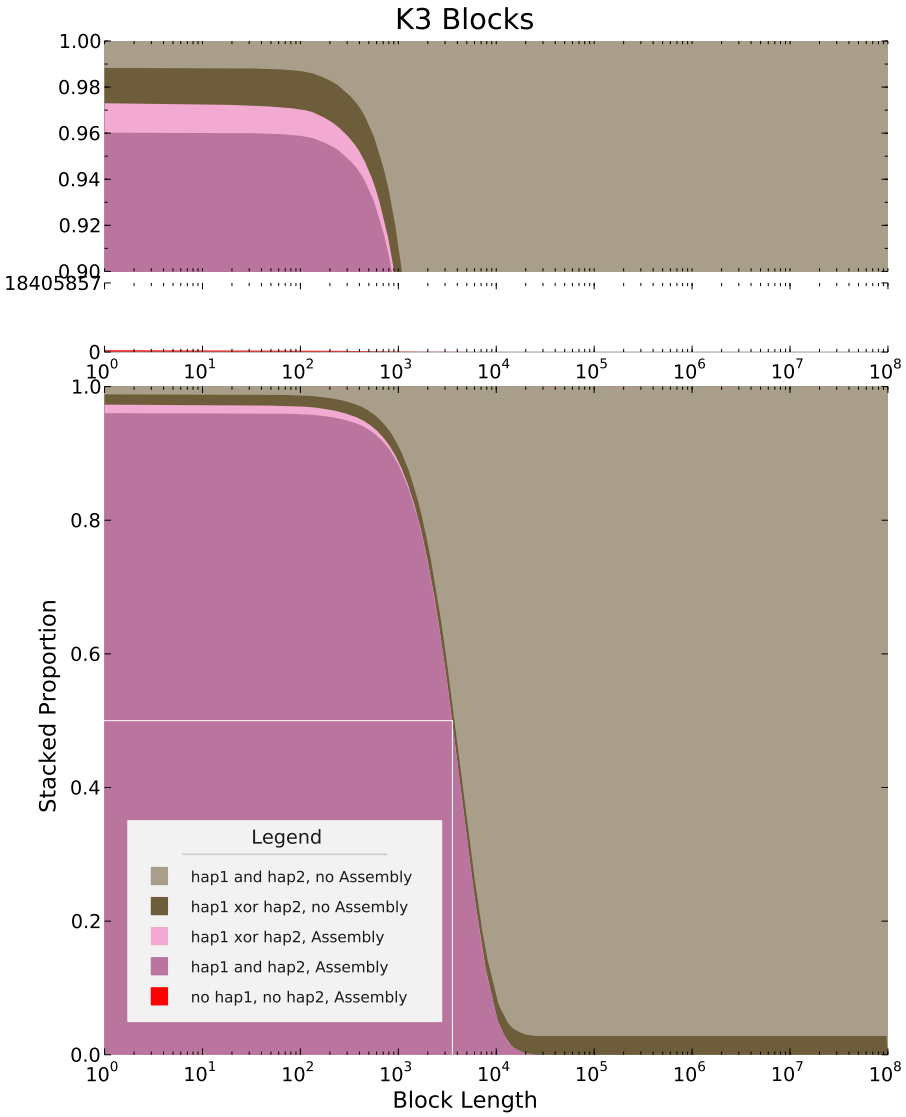


Figure 3.134: K3 blocks caption goes here.

3.2.12 L, PRICE @ deRisi Lab

Affiliation: UC San Francicso, USA
Contact: Graham Ruby
Software: **PRICE**
Number of entries: 1

ID	Total	Hap 1	Hap 2	Bac
L1	0.83722	0.83727	0.83716	0.00000

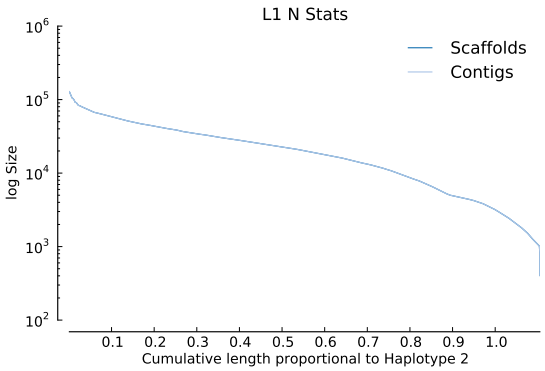
Assemblies:

L1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
N3	0.85657	0.85681	0.85634	0.00000
L1	0.83722	0.83727	0.83716	0.00000
N2	0.79226	0.79257	0.79195	0.00000

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	14,822	404	1,826.25	3,810	8,386.87	8,948.75	128,347	11,927.01	124,310,199
Contigs	14,822	404	1,826.25	3,810	8,386.87	8,948.75	128,347	11,927.01	124,310,199

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	98,281,517 – 99,262,240	80,820,755 – 81,581,285	161,619,208.0 – 163,101,680.0	11,151 – 30,441
Heterozygous	387,889 – 396,512	317,518 – 324,673	632,856.0 – 646,458.0	1,090 – 1,444
Indel	1,390,045 – 1,789,433	558,502 – 761,013	1,114,150.0 – 1,504,468.0	1,427 – 8,778

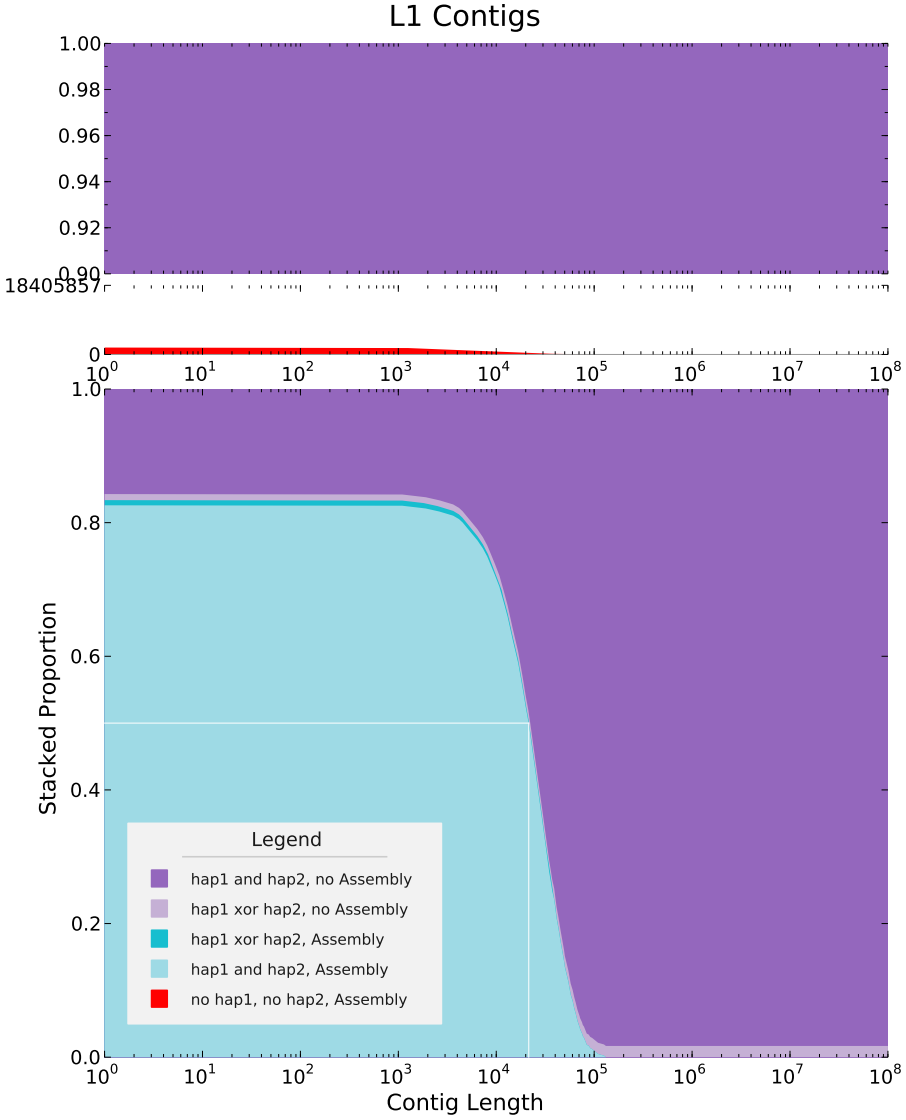


Figure 3.135: L1 contigs caption goes here.

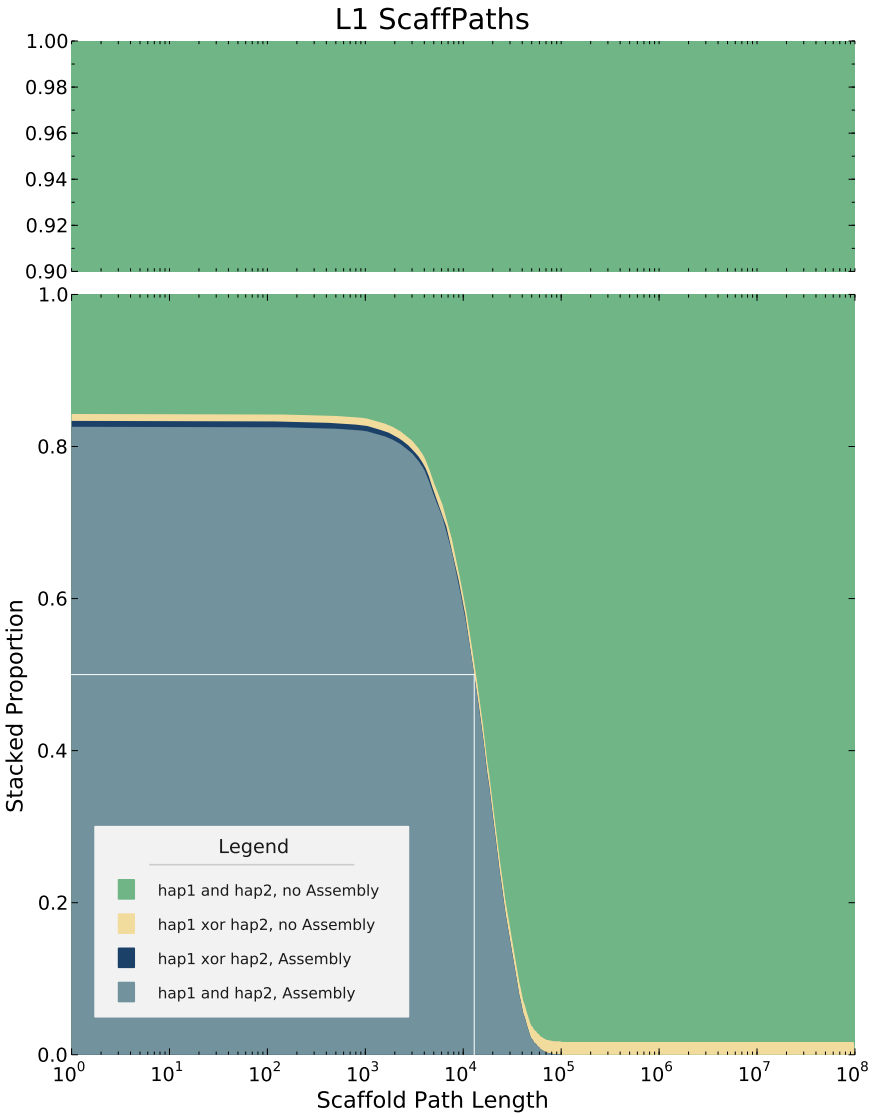


Figure 3.136: L1 scaffolds caption goes here.

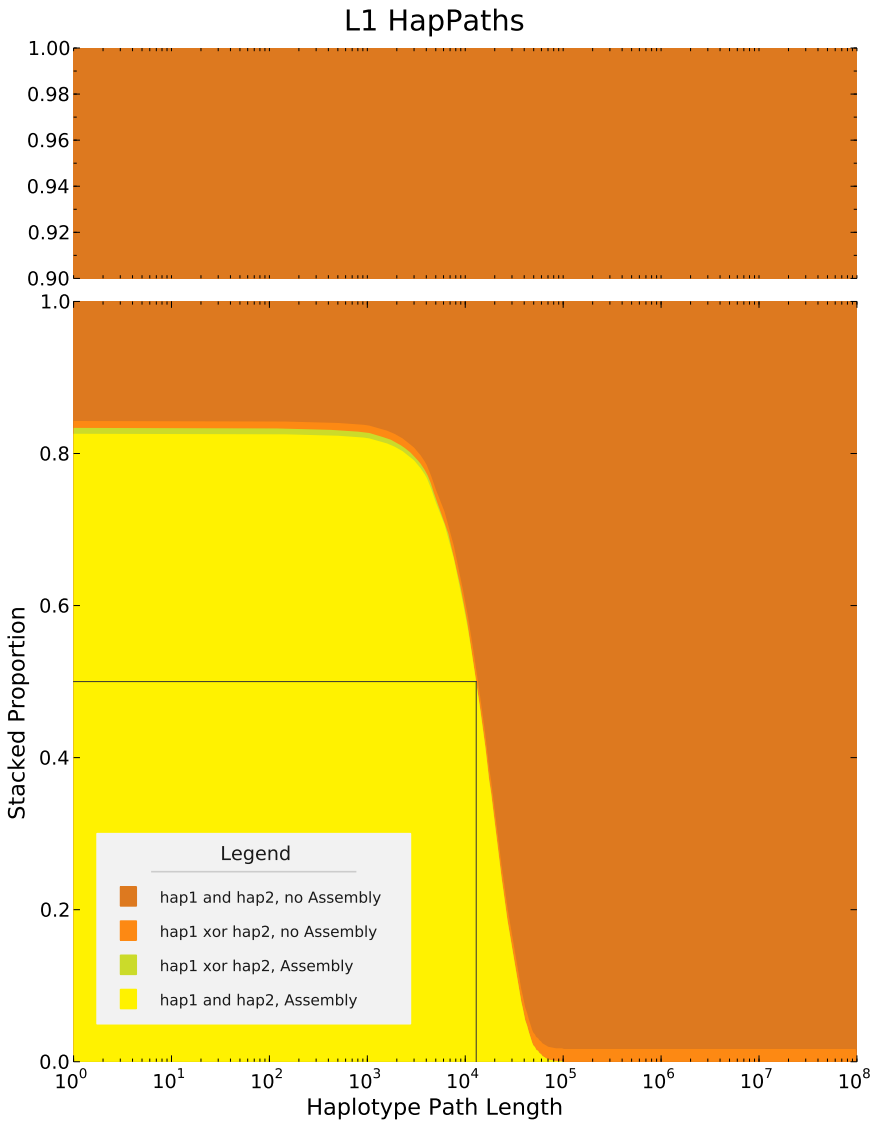


Figure 3.137: L1 hapPaths caption goes here.

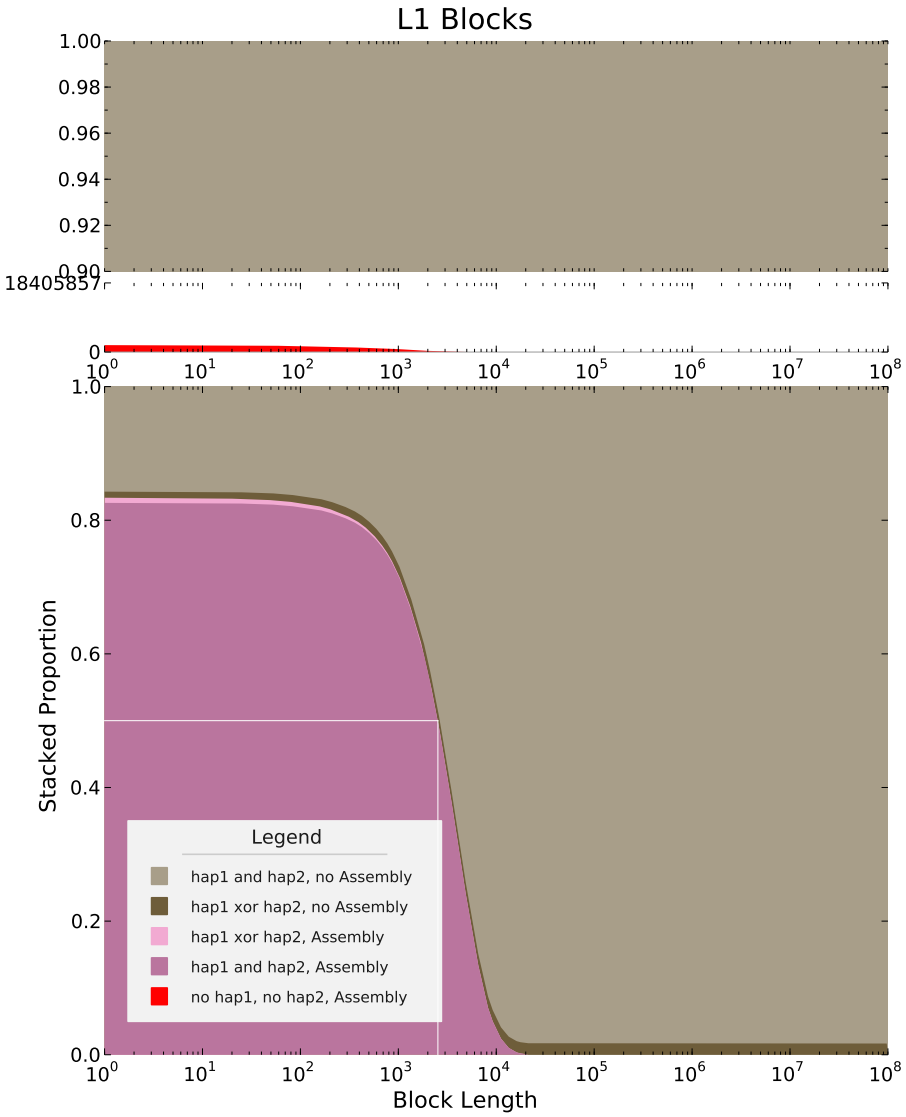


Figure 3.138: L1 blocks caption goes here.

3.2.13 M, Softberry

Affiliation: Royal Holloway, University of London, UK
Contact: Victor Solovyev
Software: **OligoZip**
Number of entries: 5

ID	Total	Hap 1	Hap 2	Bac
M3	0.98674	0.98685	0.98662	0.99977
M5	0.98667	0.98679	0.98655	0.99969
M2	0.98534	0.98552	0.98516	0.99969
M1	0.98499	0.98522	0.98477	0.99967
M4	0.97031	0.97047	0.97014	0.99718

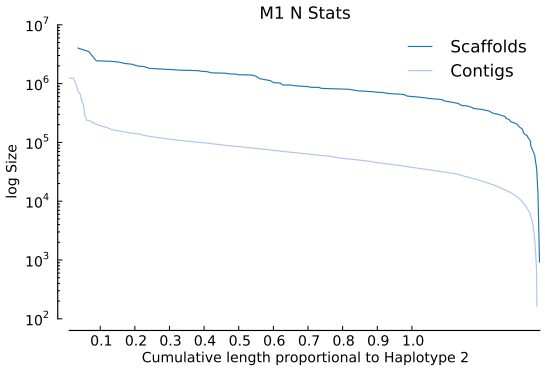
Assemblies:

M1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
M2	0.98534	0.98552	0.98516	0.99969
M1	0.98499	0.98522	0.98477	0.99967
I2	0.98467	0.98511	0.98424	0.99857

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	657	892	1,014.00	1,418	234,536.90	279,743.00	4,022,579	475,732.45	154,090,743
Contigs	4,477	157	5,835.00	21,167	34,208.10	46,473.00	1,239,957	49,902.51	153,149,674

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	79,070,669 – 79,512,246	78,367,042 – 78,759,434	156,732,786.0 – 157,515,016.0	491 – 1,314
Heterozygous	311,166 – 317,542	307,843 – 313,823	615,678.0 – 627,622.0	2 – 6
Indel	2,746,787 – 3,080,879	1,070,026 – 1,224,885	2,136,464.0 – 2,445,266.0	1,791 – 2,143

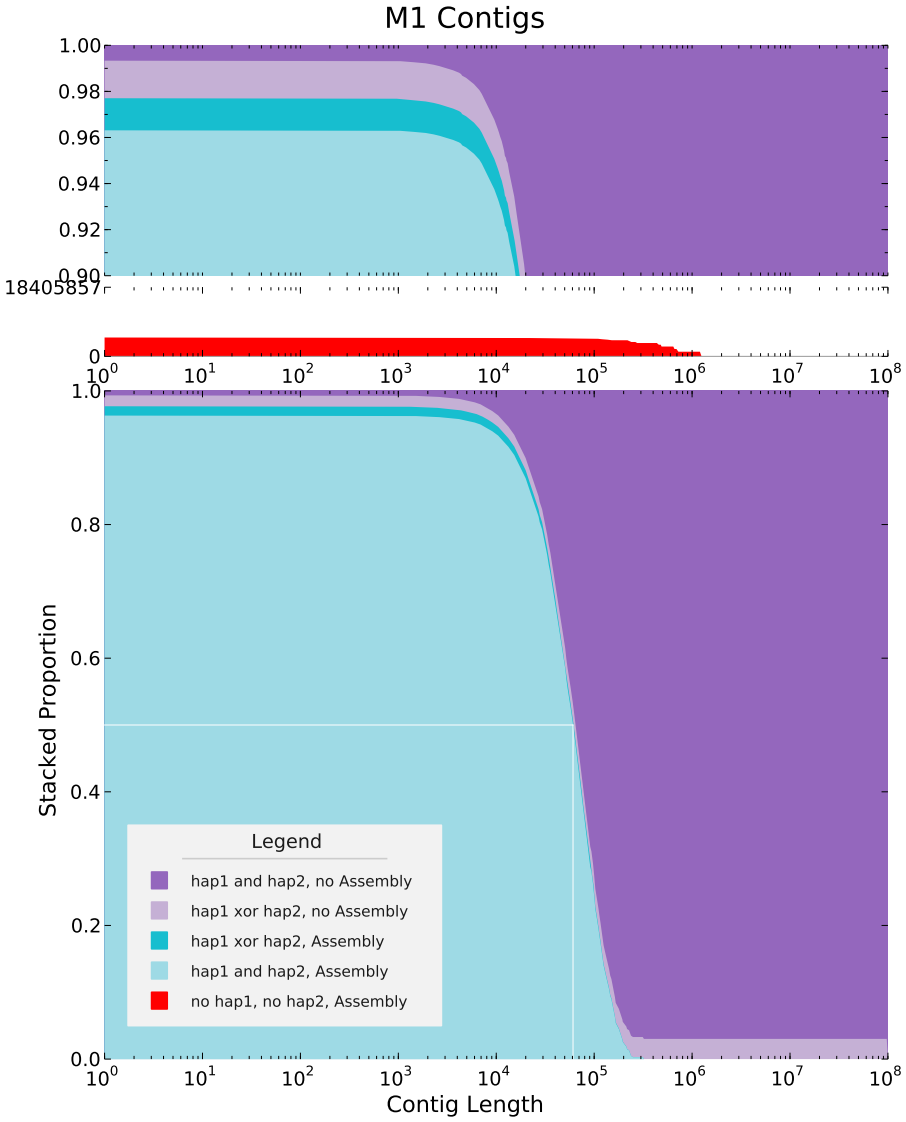


Figure 3.139: M1 contigs caption goes here.

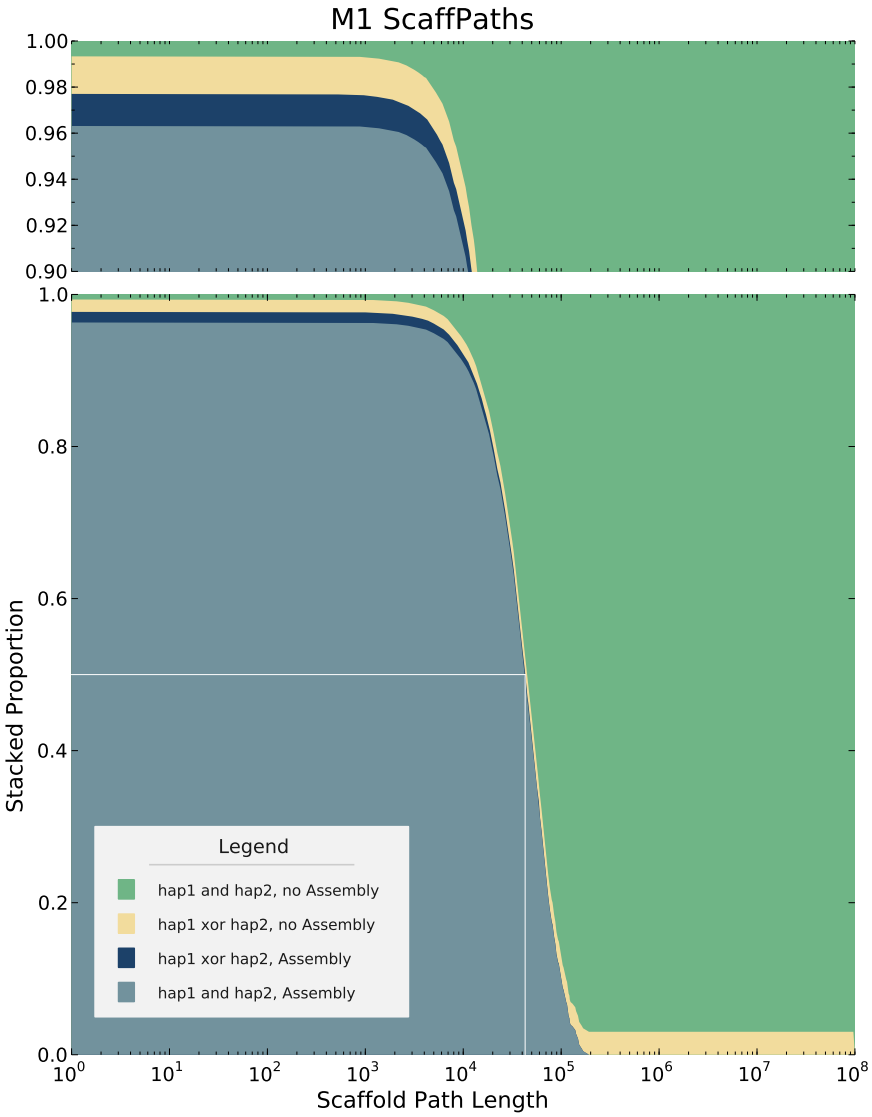


Figure 3.140: M1 scaffolds caption goes here.

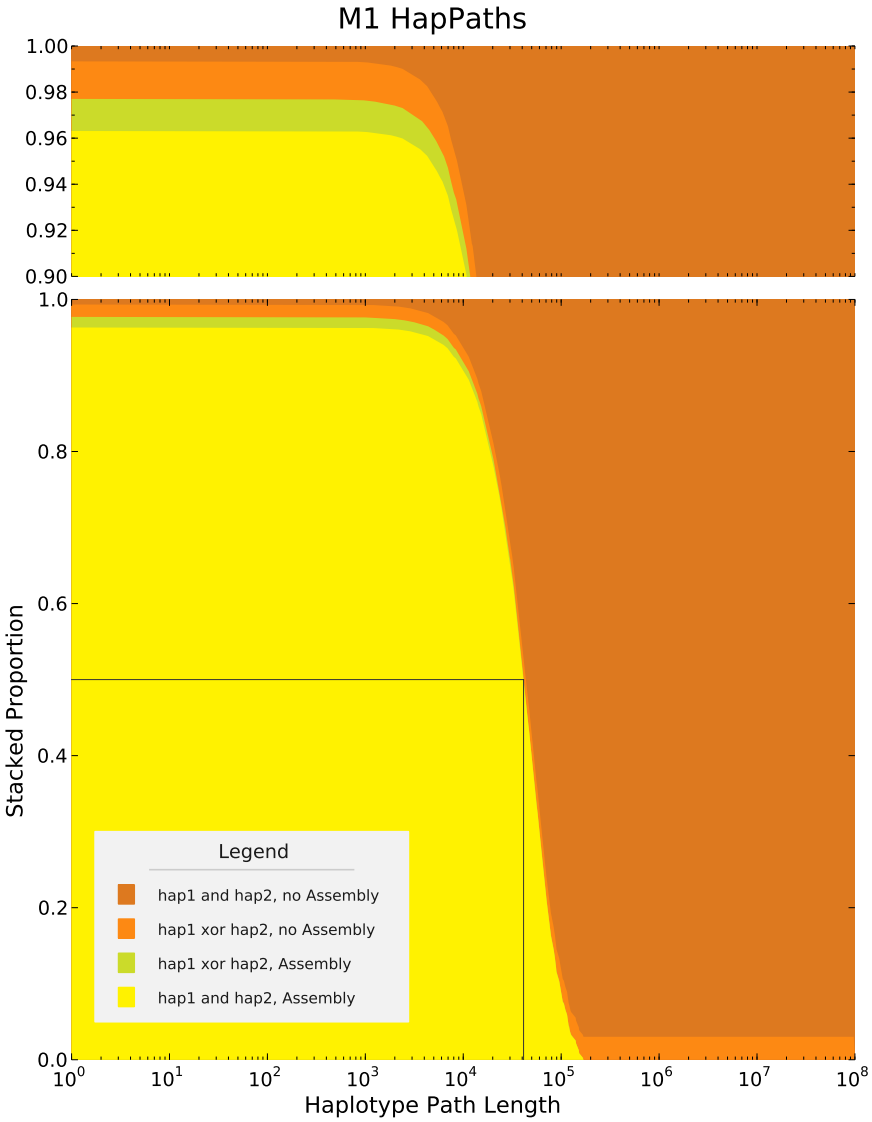


Figure 3.141: M1 hapPaths caption goes here.

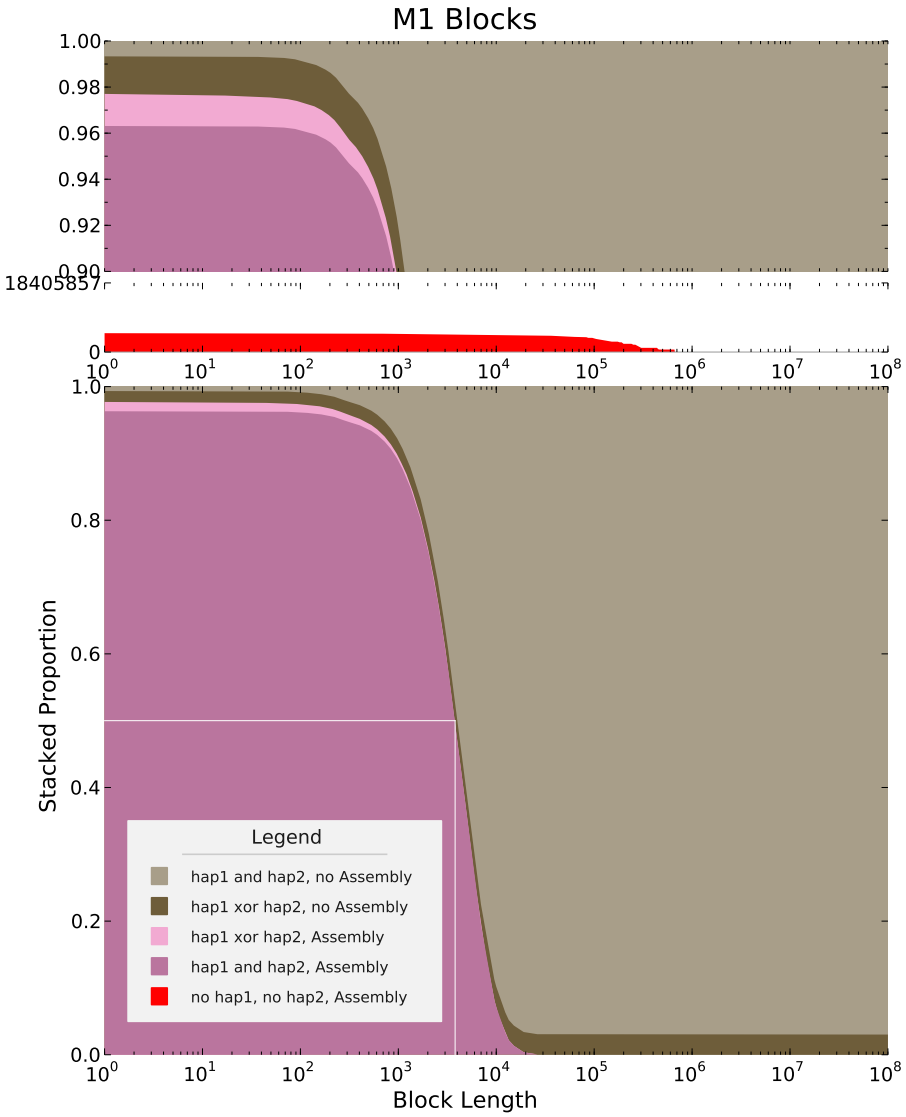


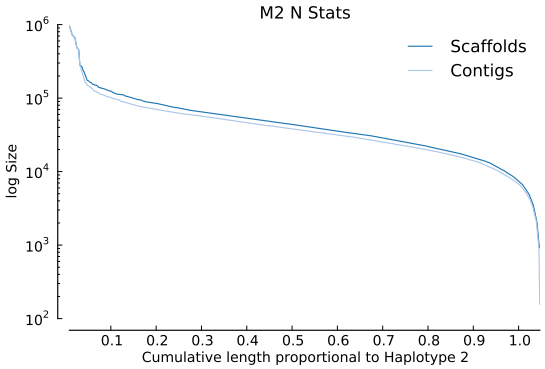
Figure 3.142: M1 blocks caption goes here.

M2

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
B2	0.98568	0.98600	0.98535	0.99892
M2	0.98534	0.98552	0.98516	0.99969
M1	0.98499	0.98522	0.98477	0.99967

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	5,000	892	5,488.50	14,900	23,559.50	31,230.75	959,012	33,210.96	117,797,484
Contigs	5,780	157	4,439.50	13,167	20,372.40	27,442.00	959,012	29,537.52	117,752,457

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,034,382 – 108,626,598	107,329,830 – 107,870,867	214,658,544.0 – 215,738,488.0	521 – 1,438
Heterozygous	424,871 – 433,450	421,552 – 429,718	843,098.0 – 859,416.0	3 – 9
Indel	2,997,904 – 3,390,024	1,401,048 – 1,614,139	2,797,478.0 – 3,222,540.0	2,307 – 2,800

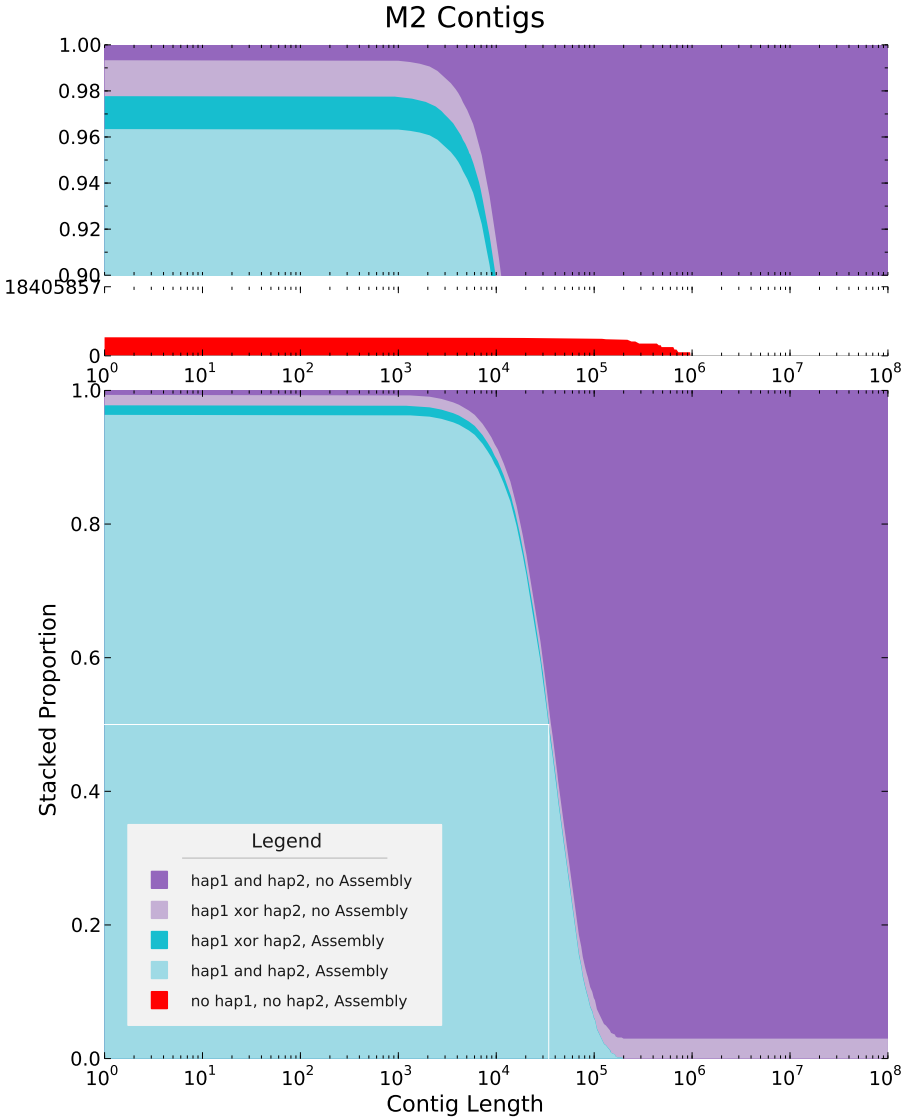


Figure 3.143: M2 contigs caption goes here.

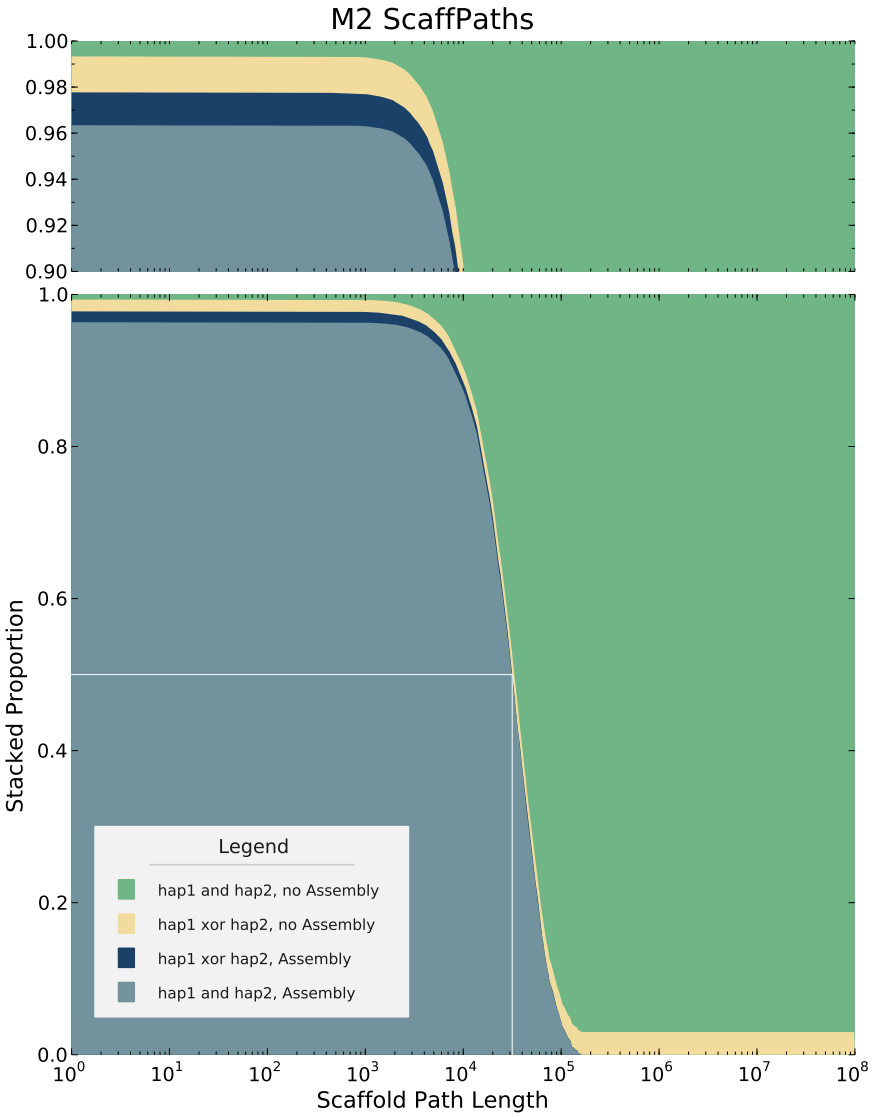


Figure 3.144: M2 scaffolds caption goes here.

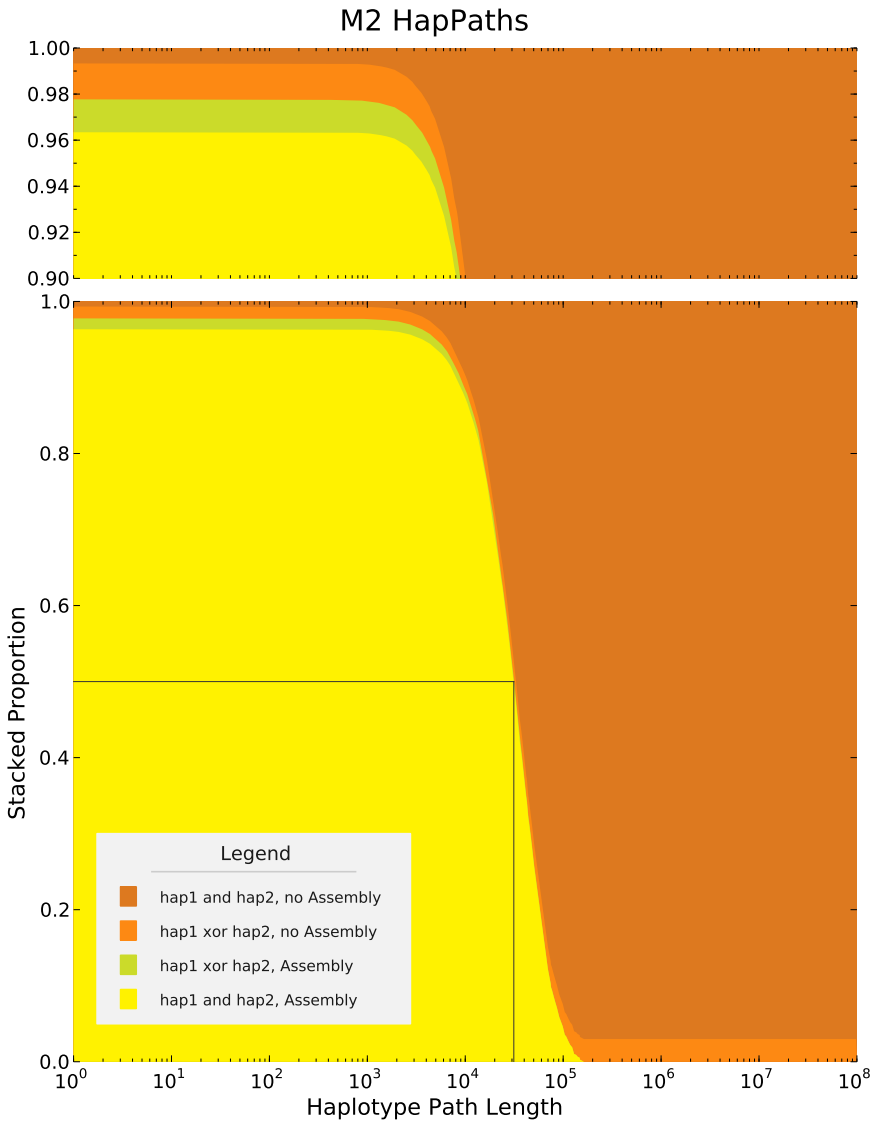


Figure 3.145: M2 hapPaths caption goes here.

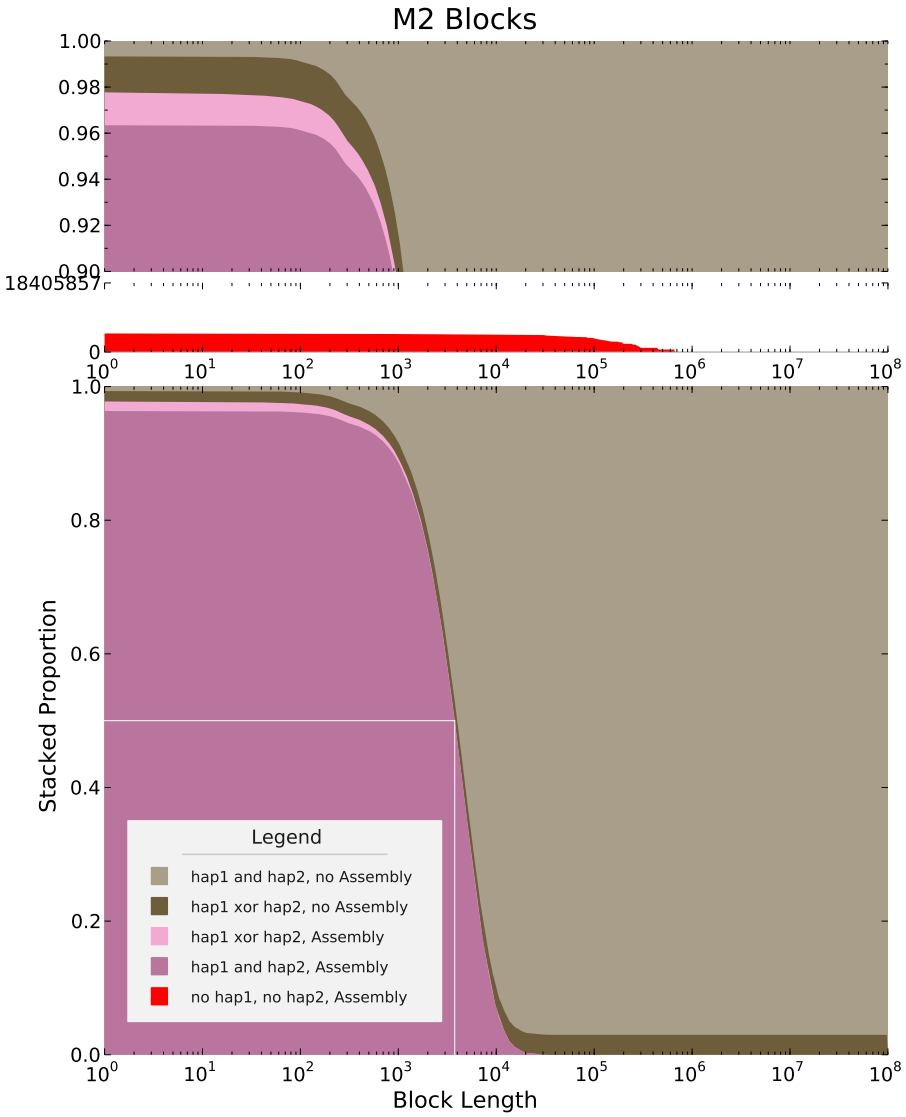


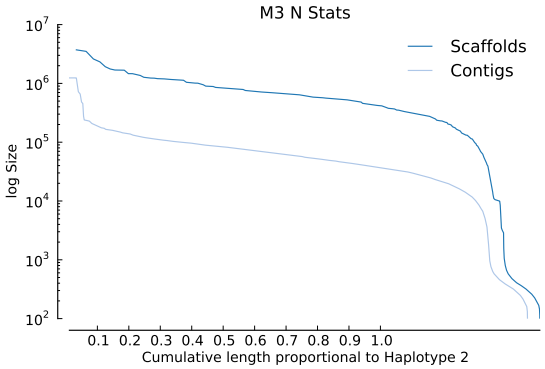
Figure 3.146: M2 blocks caption goes here.

M3

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
F5	0.98691	0.98727	0.98653	0.99934
M3	0.98674	0.98685	0.98662	0.99977
F3	0.98671	0.98696	0.98648	0.99927

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	40,344	101	226.00	306	4,203.08	400.00	3,720,691	58,324.61	169,568,979
Contigs	45,200	101	236.00	324	3,653.81	439.00	1,240,157	18,810.80	165,152,290

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	71,551,605 – 71,969,021	70,904,442 – 71,275,640	141,807,836.0 – 142,547,308.0	442 – 1,568
Heterozygous	277,102 – 281,107	274,023 – 277,682	548,038.0 – 555,340.0	3 – 8
Indel	2,624,712 – 2,995,427	1,074,161 – 1,408,975	2,144,914.0 – 2,813,152.0	1,704 – 2,266

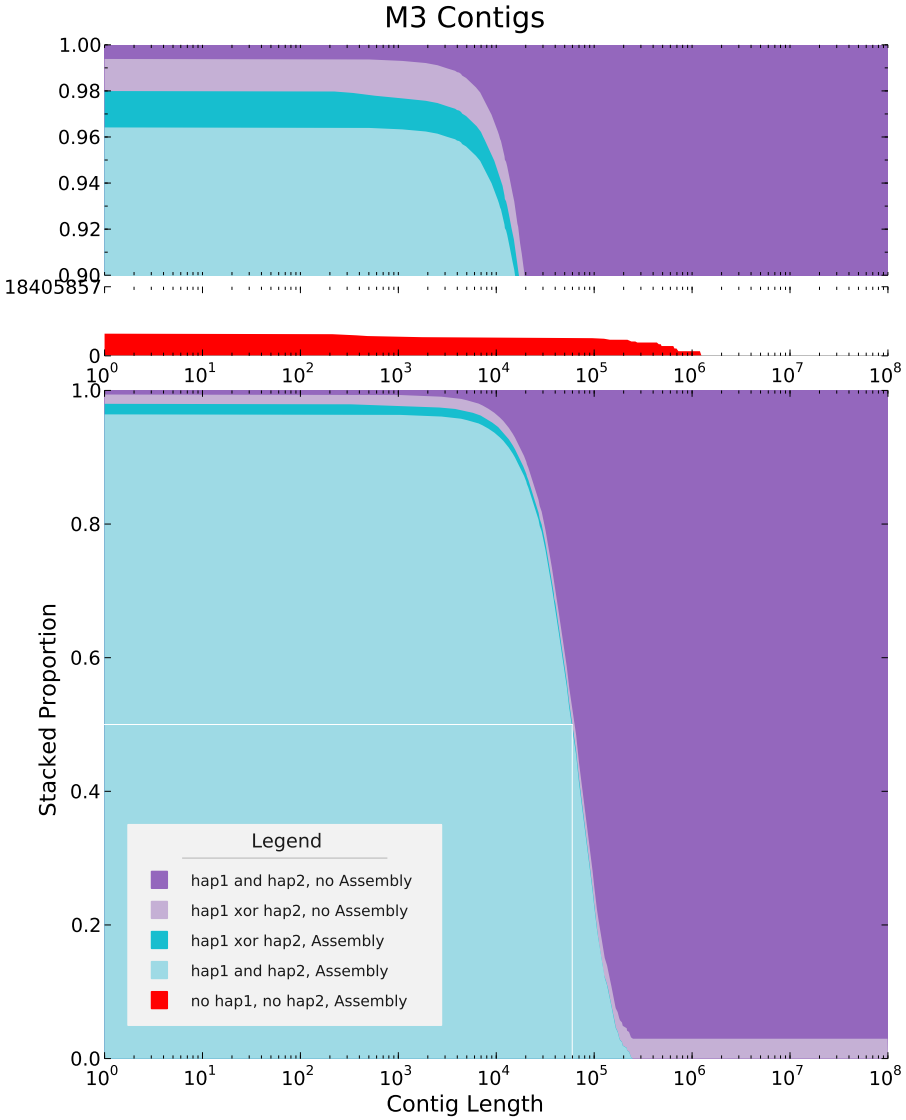


Figure 3.147: M3 contigs caption goes here.

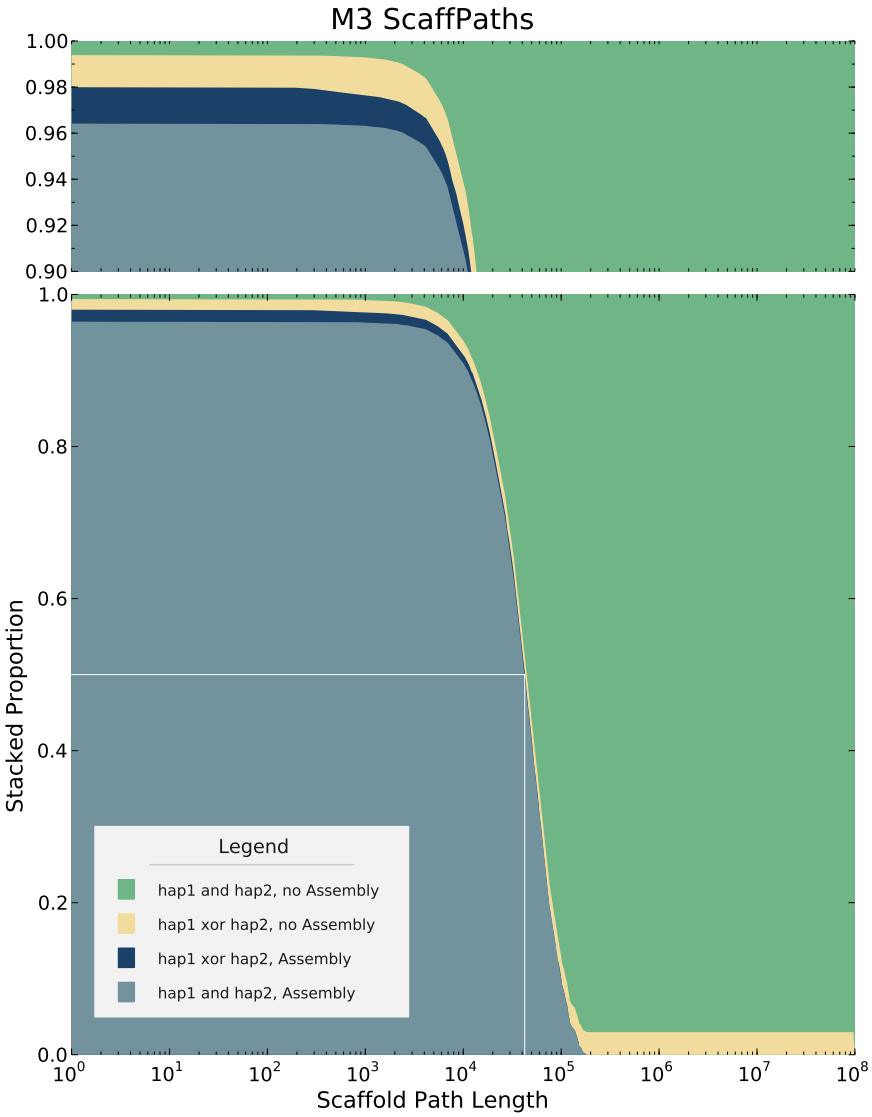


Figure 3.148: M3 scaffolds caption goes here.

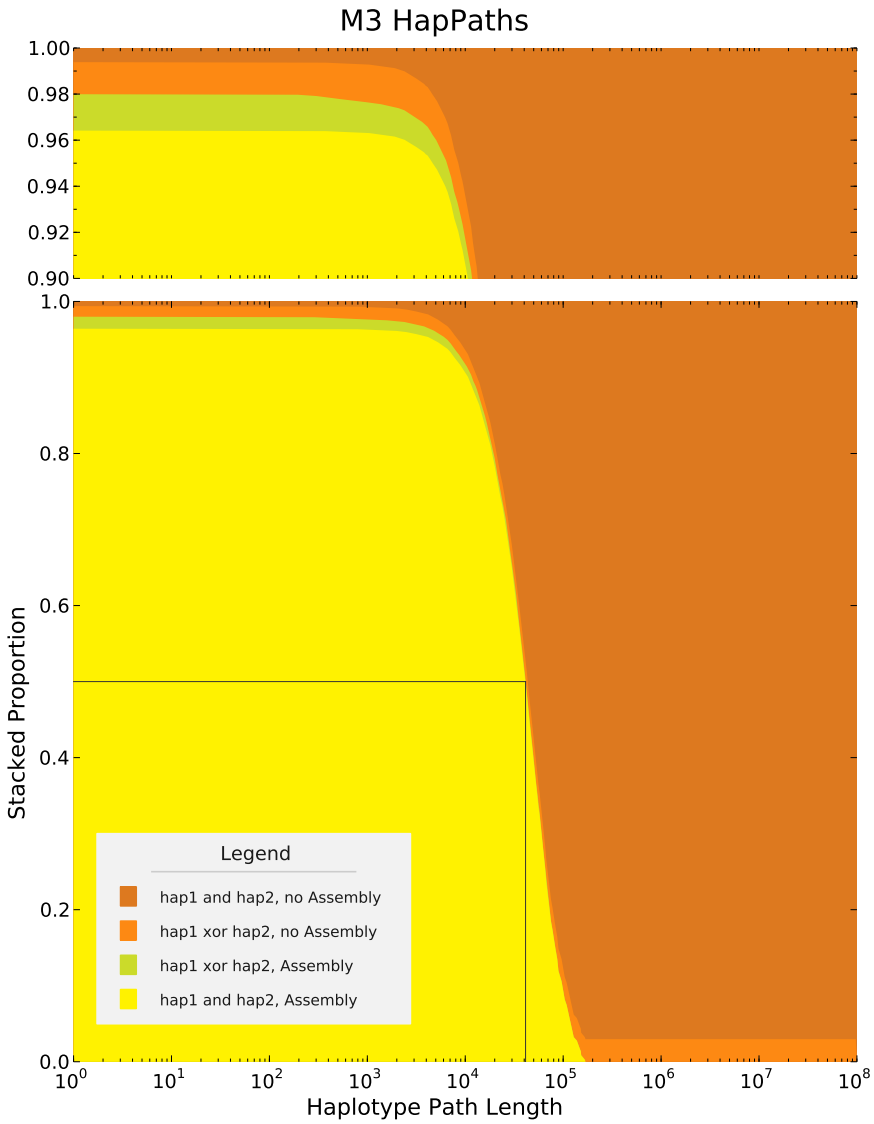


Figure 3.149: M3 hapPaths caption goes here.

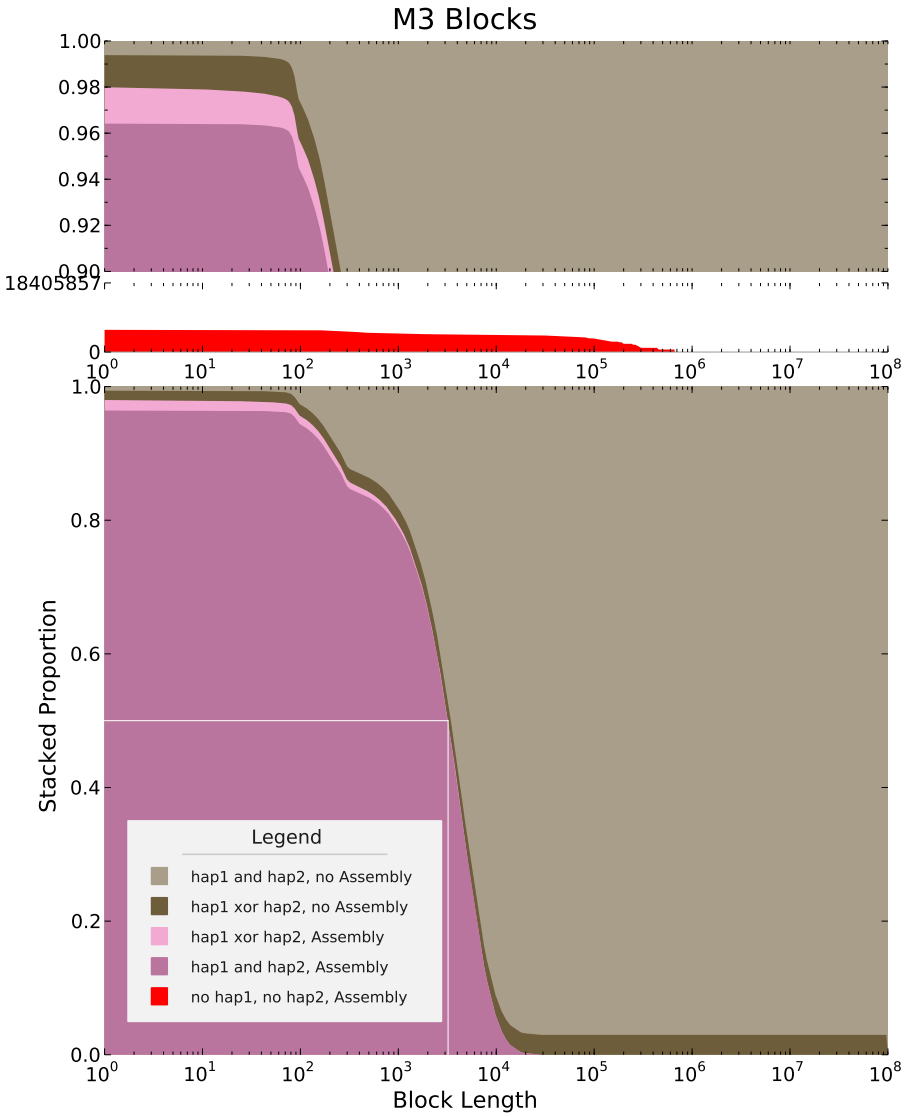


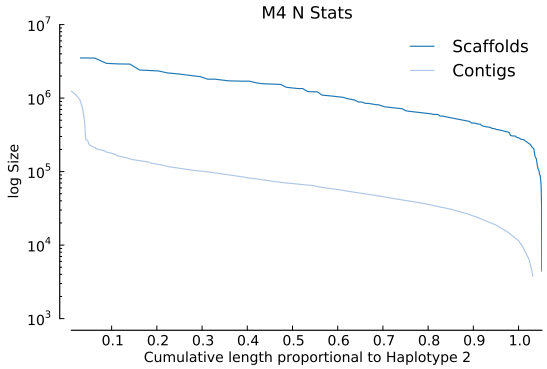
Figure 3.150: M3 blocks caption goes here.

M4

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
W1	0.97034	0.97048	0.97023	0.99825
M4	0.97031	0.97047	0.97014	0.99718
W7	0.96984	0.97006	0.96961	0.99806

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	158	4,447	280,708.75	513,458	748,859.06	957,382.00	3,518,296	715,708.11	118,319,732
Contigs	2,672	3,663	14,400.75	30,801	43,453.11	57,323.75	1,239,957	54,475.36	116,106,704

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,784,300 – 109,379,247	106,407,542 – 106,944,439	212,813,426.0 – 213,884,110.0	609 – 1,484
Heterozygous	428,115 – 436,978	417,966 – 426,307	835,920.0 – 852,580.0	3 – 9
Indel	2,910,014 – 3,294,504	1,349,054 – 1,542,590	2,693,794.0 – 3,080,006.0	2,155 – 2,514

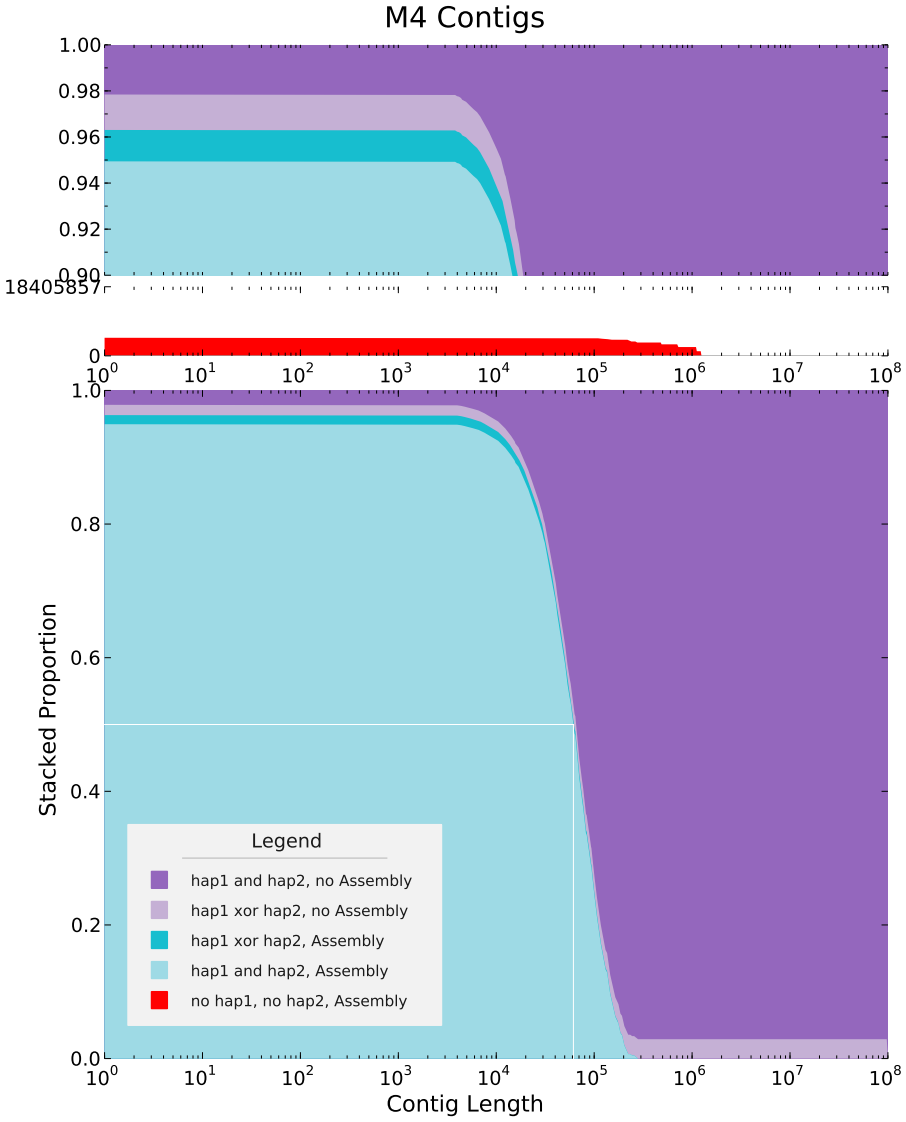


Figure 3.151: M4 contigs caption goes here.

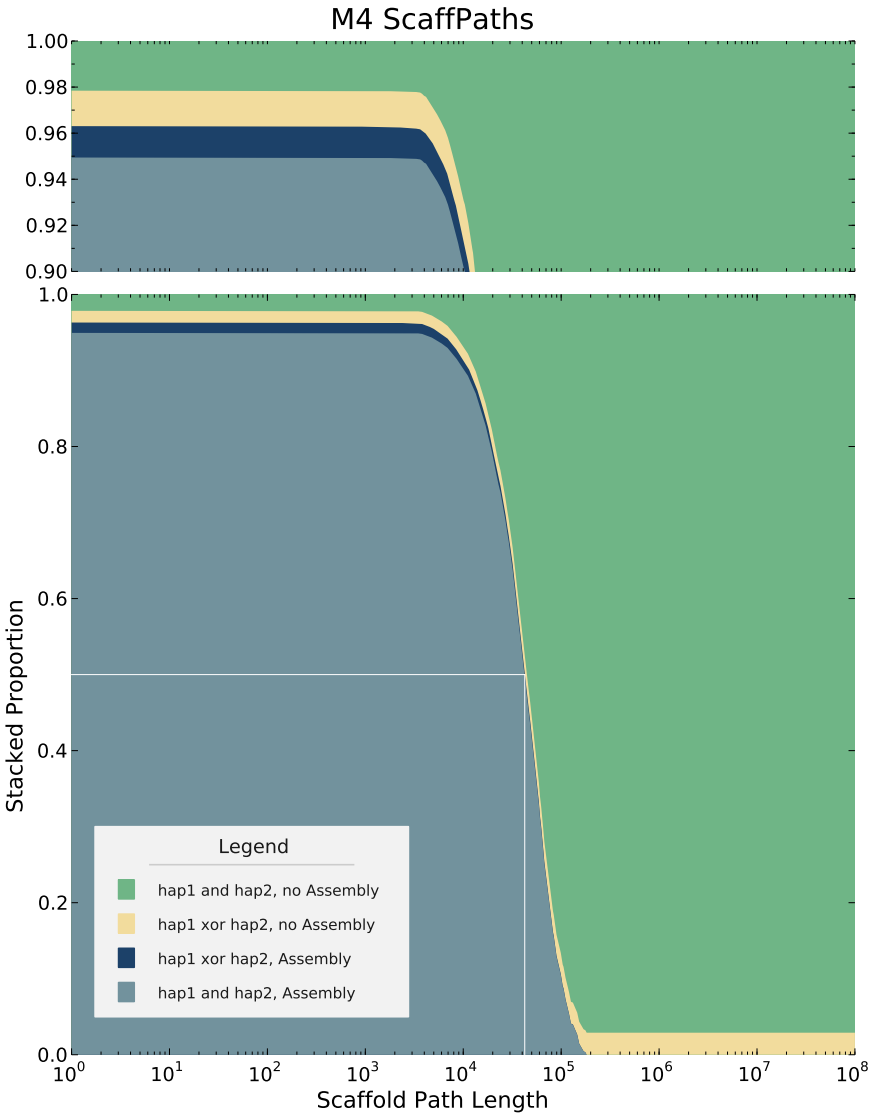


Figure 3.152: M4 scaffolds caption goes here.

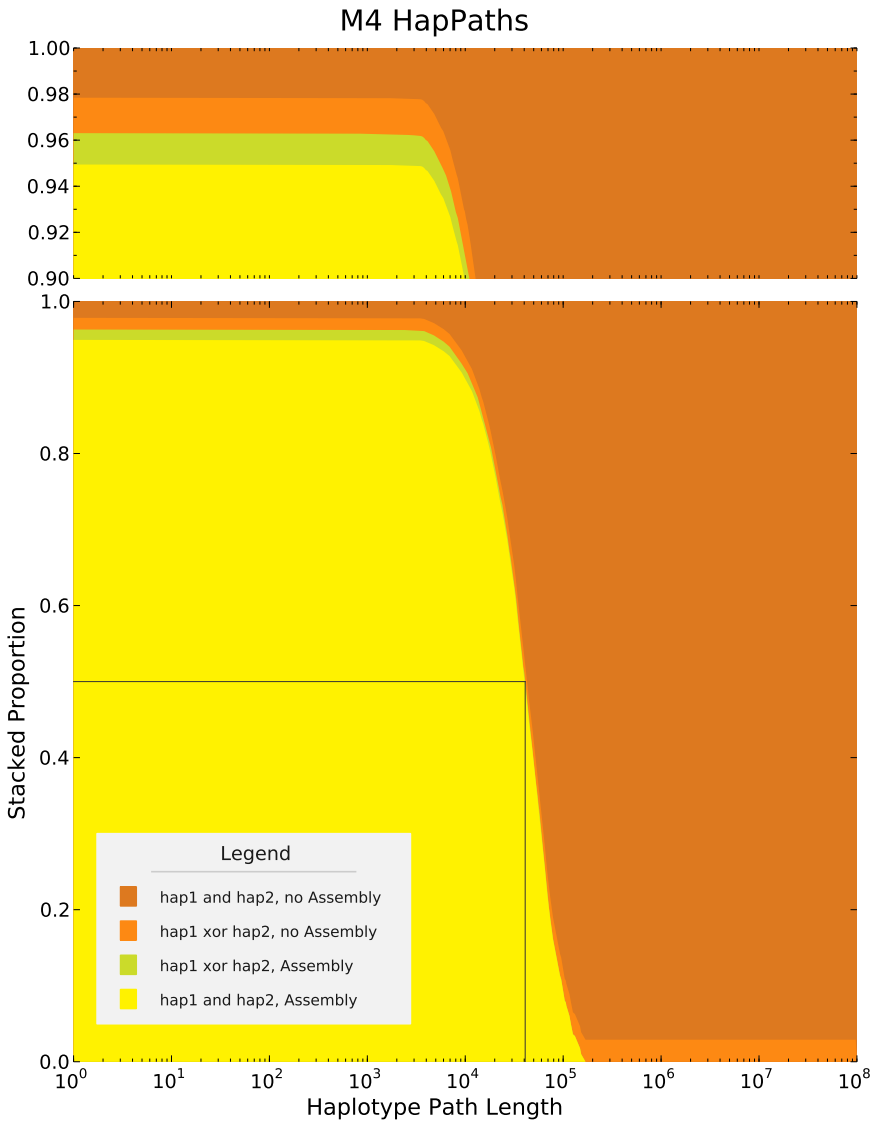


Figure 3.153: M4 hapPaths caption goes here.

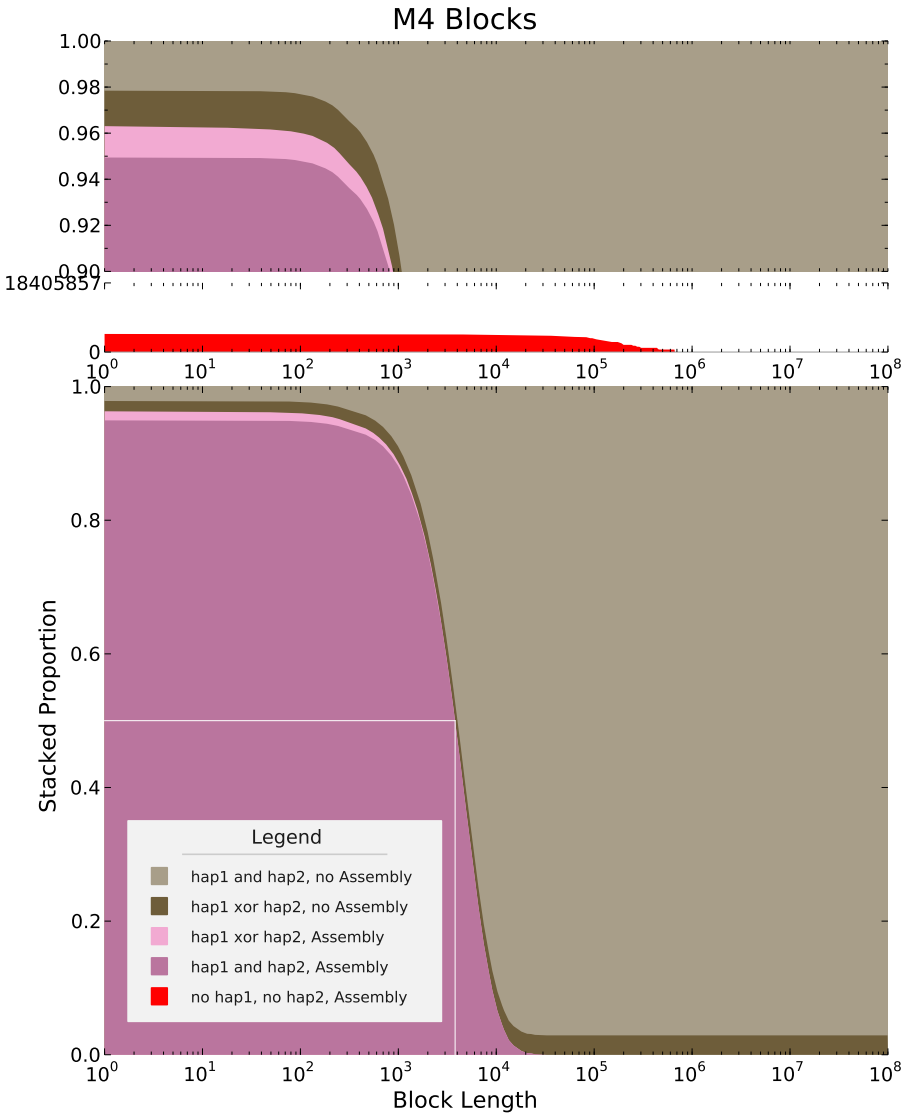


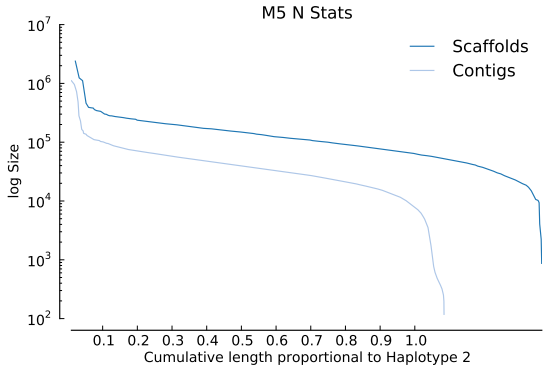
Figure 3.154: M4 blocks caption goes here.

M5

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
F3	0.98671	0.98696	0.98648	0.99927
M5	0.98667	0.98679	0.98655	0.99969
F4	0.98649	0.98675	0.98624	0.99928

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	2,500	832	16,394.50	38,607	61,494.78	81,551.50	2,390,989	85,382.22	153,736,947
Contigs	13,998	118	356.25	524	8,716.73	8,565.75	1,111,891	22,488.26	122,016,803

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	104,581,692 – 105,149,319	103,977,385 – 104,499,826	207,952,908.0 – 208,993,426.0	707 – 2,164
Heterozygous	410,098 – 417,440	407,276 – 414,269	814,540.0 – 828,518.0	3 – 5
Indel	3,002,862 – 3,407,750	1,433,840 – 1,708,030	2,862,804.0 – 3,409,238.0	2,433 – 3,135

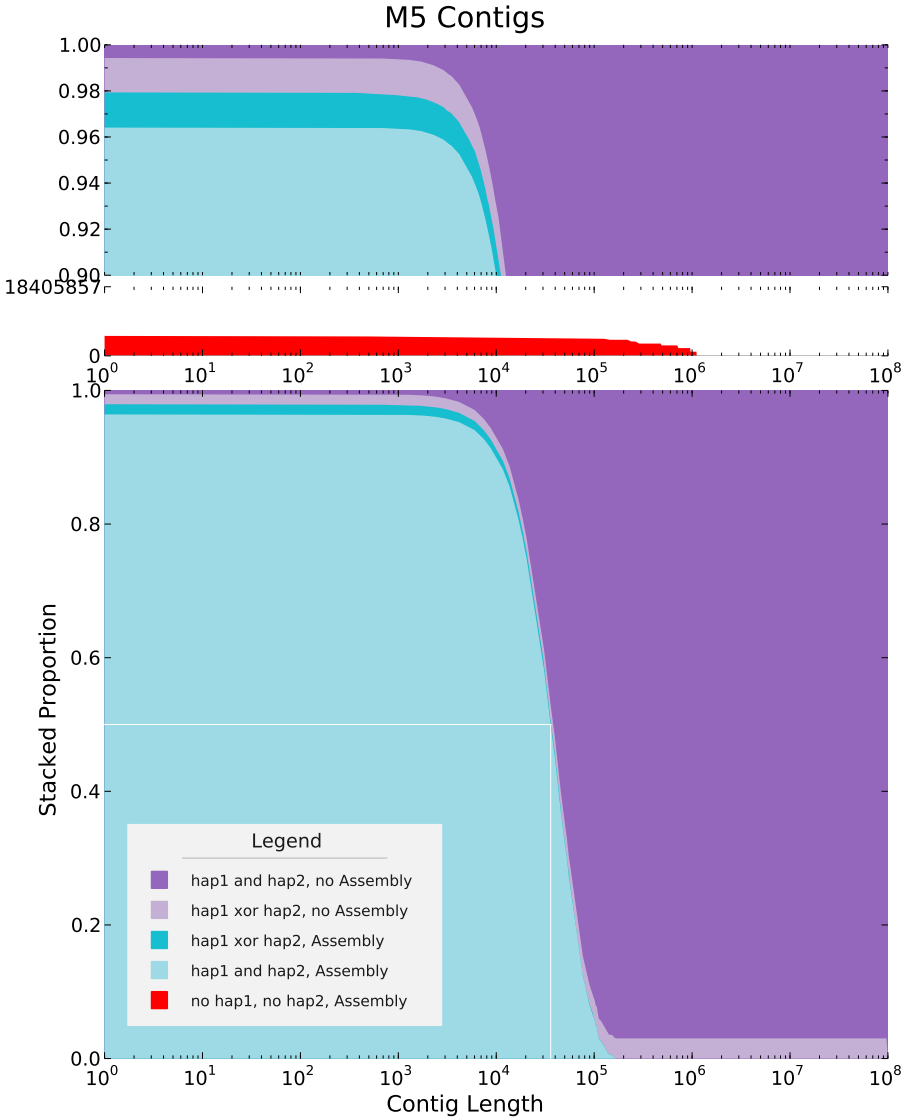


Figure 3.155: M5 contigs caption goes here.

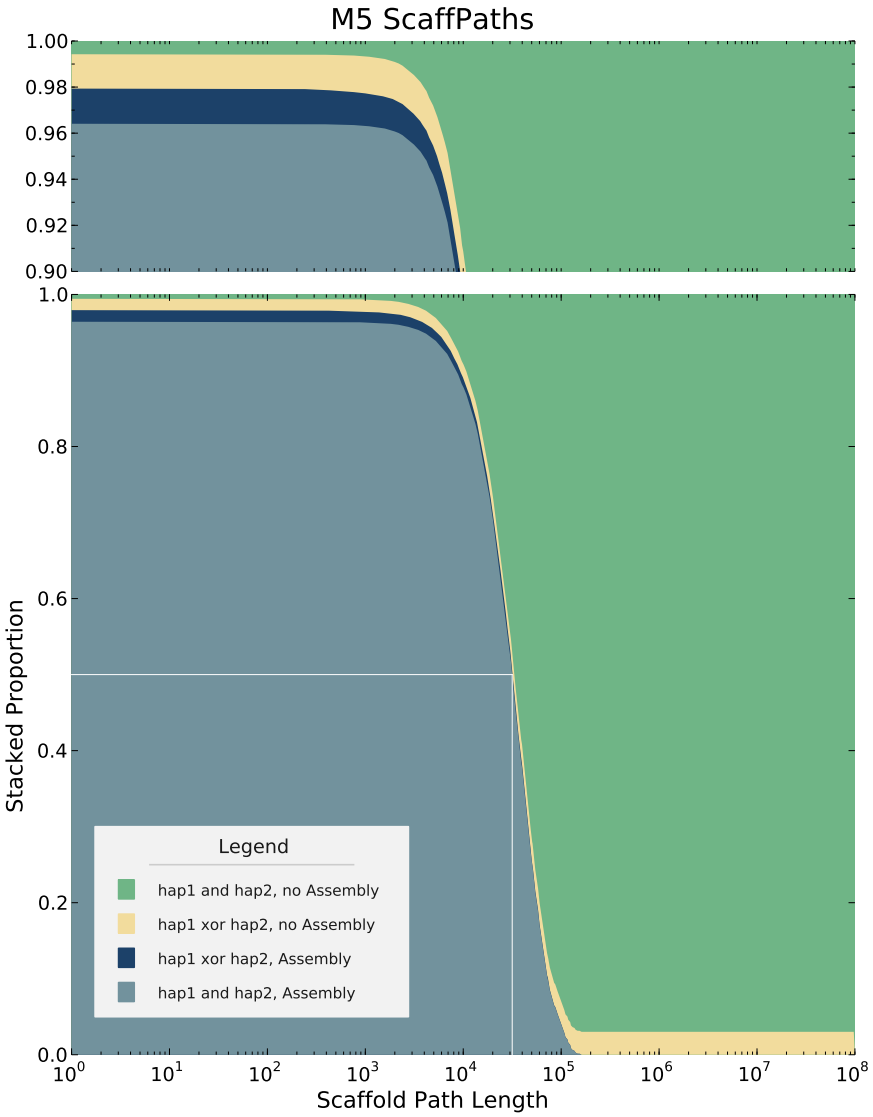


Figure 3.156: M5 scaffolds caption goes here.

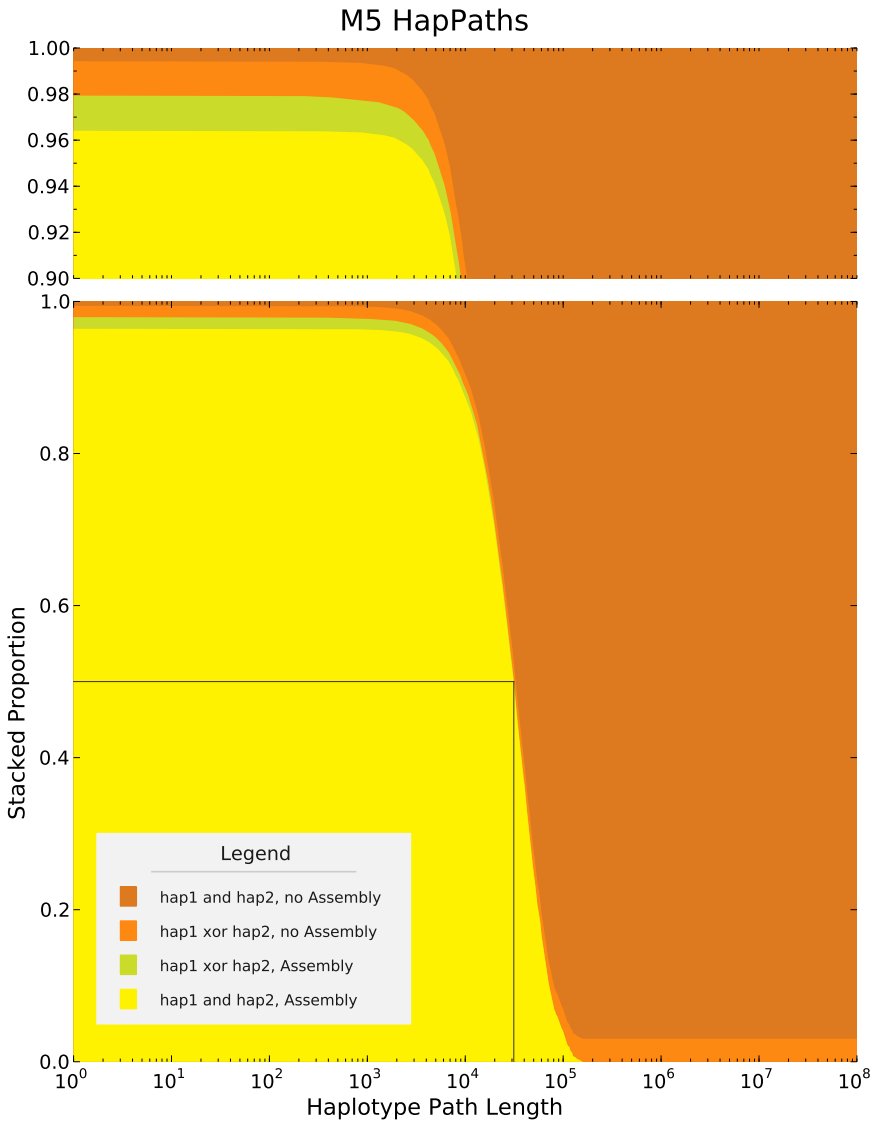


Figure 3.157: M5 hapPaths caption goes here.

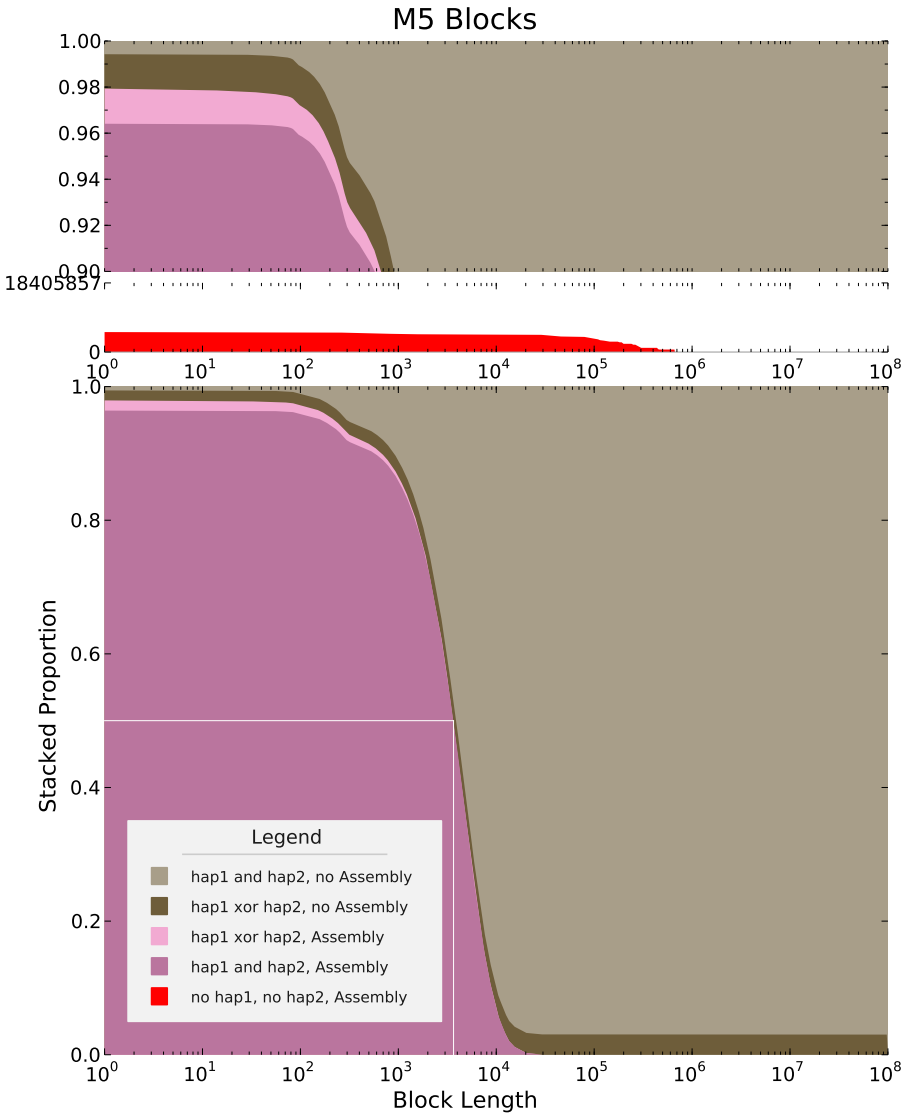


Figure 3.158: M5 blocks caption goes here.

3.2.14 N, TGAC/TSL/Oxford

Affiliation: The Genome Analysis Centre, Sainsbury Laboratory, and Wellcome Trust Centre for Human Genetics, UK
Contact: Mario Caccamo
Software: **Cortex_con_rp**
Number of entries: 3

ID	Total	Hap 1	Hap 2	Bac
N1	0.87107	0.87121	0.87094	0.00000
N3	0.85657	0.85681	0.85634	0.00000
N2	0.79226	0.79257	0.79195	0.00000

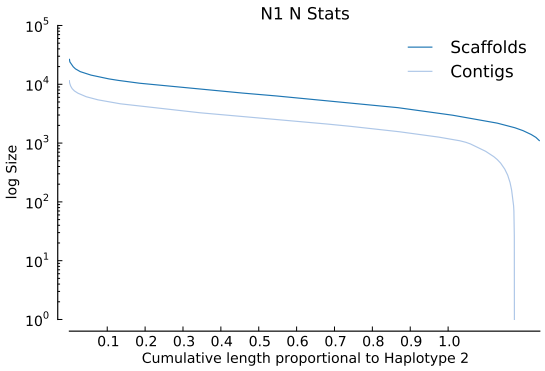
Assemblies:

N1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
I1	0.87175	0.87213	0.87138	0.99691
N1	0.87107	0.87121	0.87094	0.00000
N3	0.85657	0.85681	0.85634	0.00000

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	31,423	1,081	2,210.00	3,571	4,448.15	5,848.50	26,513	3,038.11	139,774,320
Contigs	86,428	1	455.00	1,270	1,529.91	2,198.00	11,501	1,335.38	132,226,959

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	80,315,755 – 83,700,670	68,087,614 – 70,113,684	136,124,878.0 – 140,127,048.0	3,872 – 11,050
Heterozygous	305,379 – 339,277	247,666 – 262,788	493,466.0 – 521,994.0	19 – 55
Indel	1,752,542 – 2,129,827	566,357 – 703,522	1,130,822.0 – 1,400,404.0	813 – 1,487

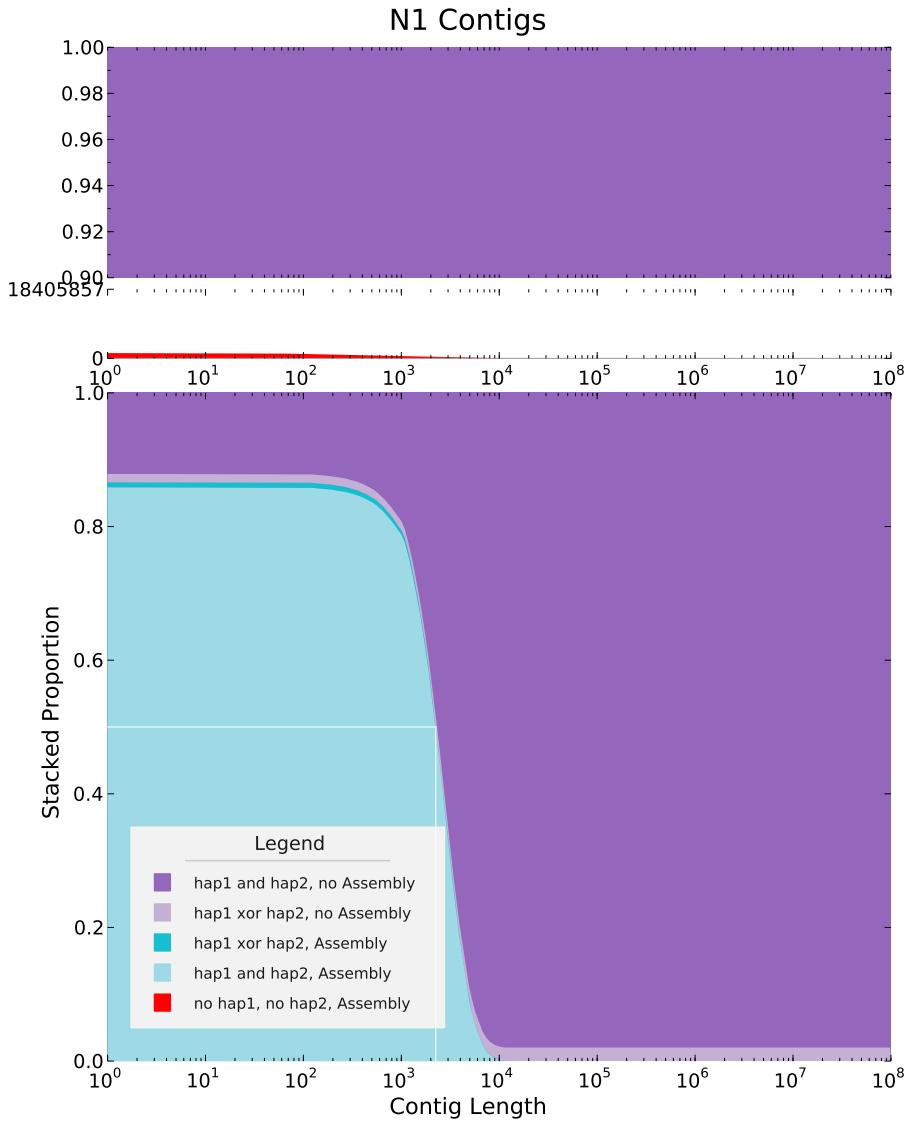


Figure 3.159: N1 contigs caption goes here.

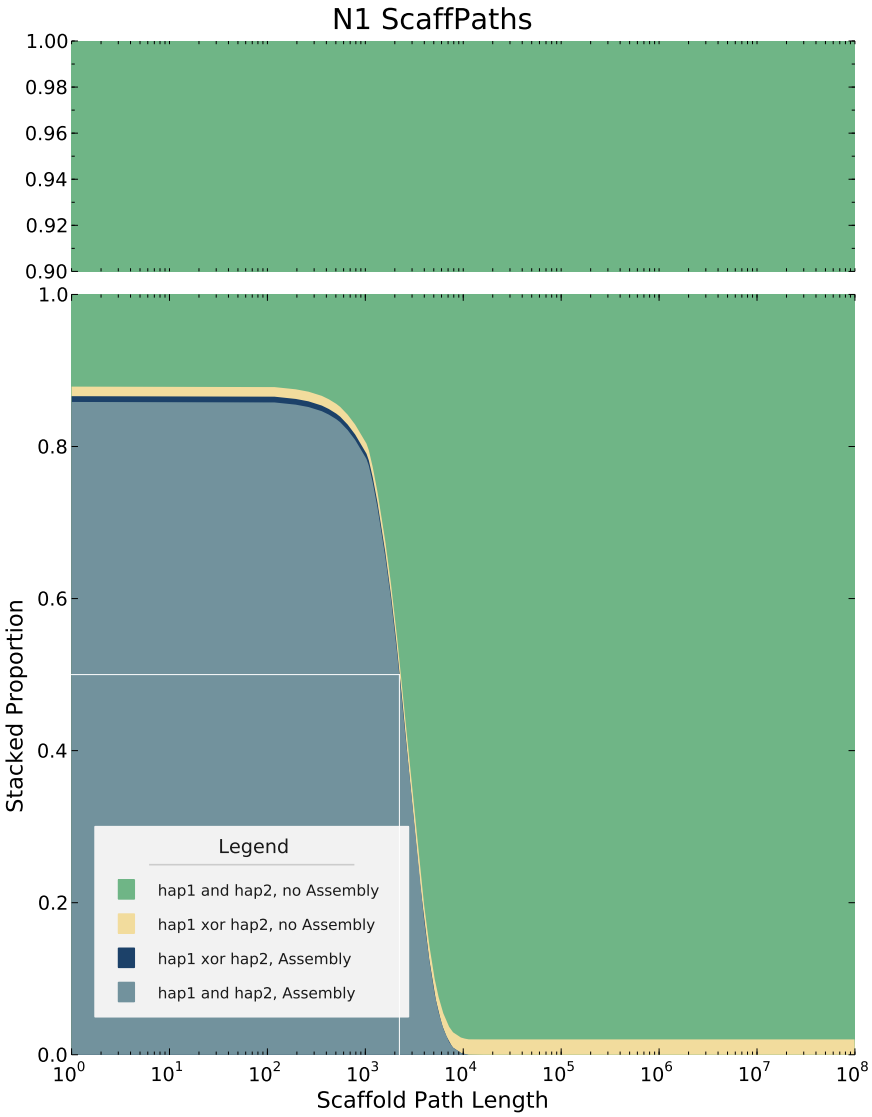


Figure 3.160: N1 scaffolds caption goes here.

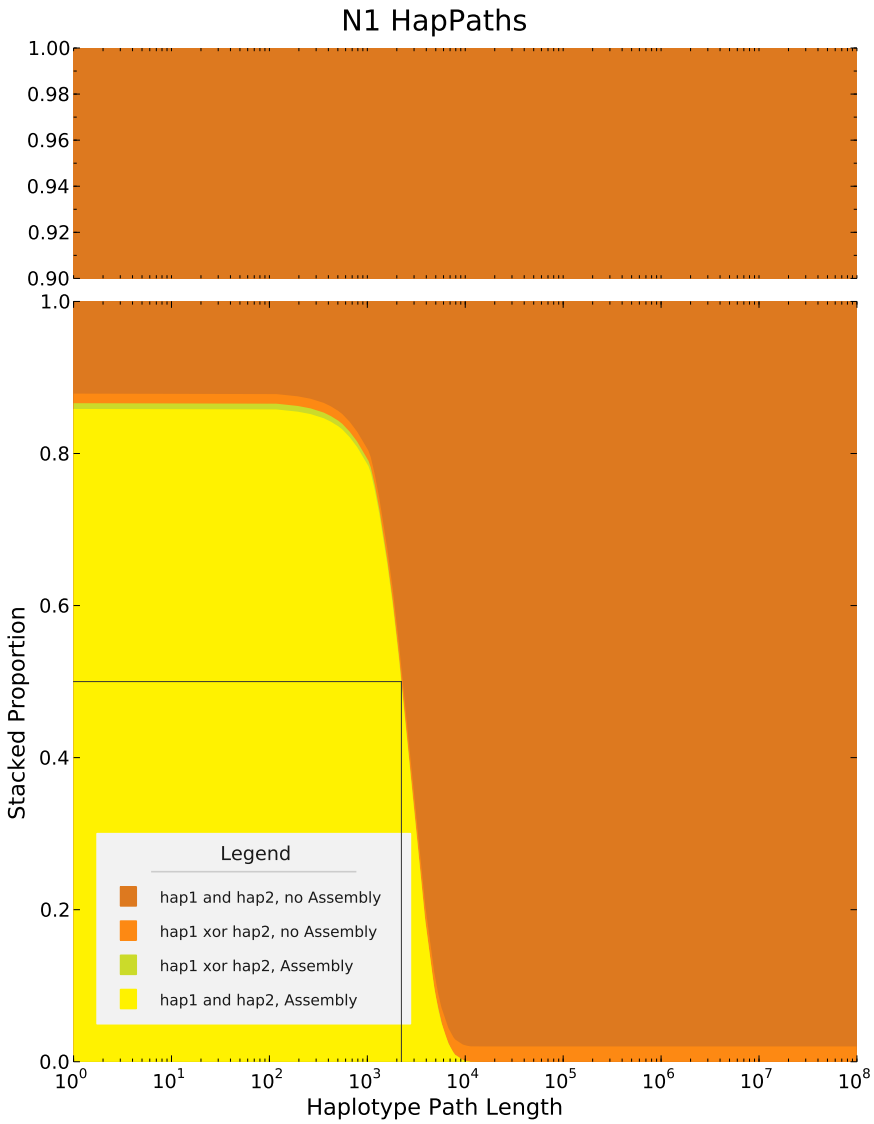


Figure 3.161: N1 hapPaths caption goes here.

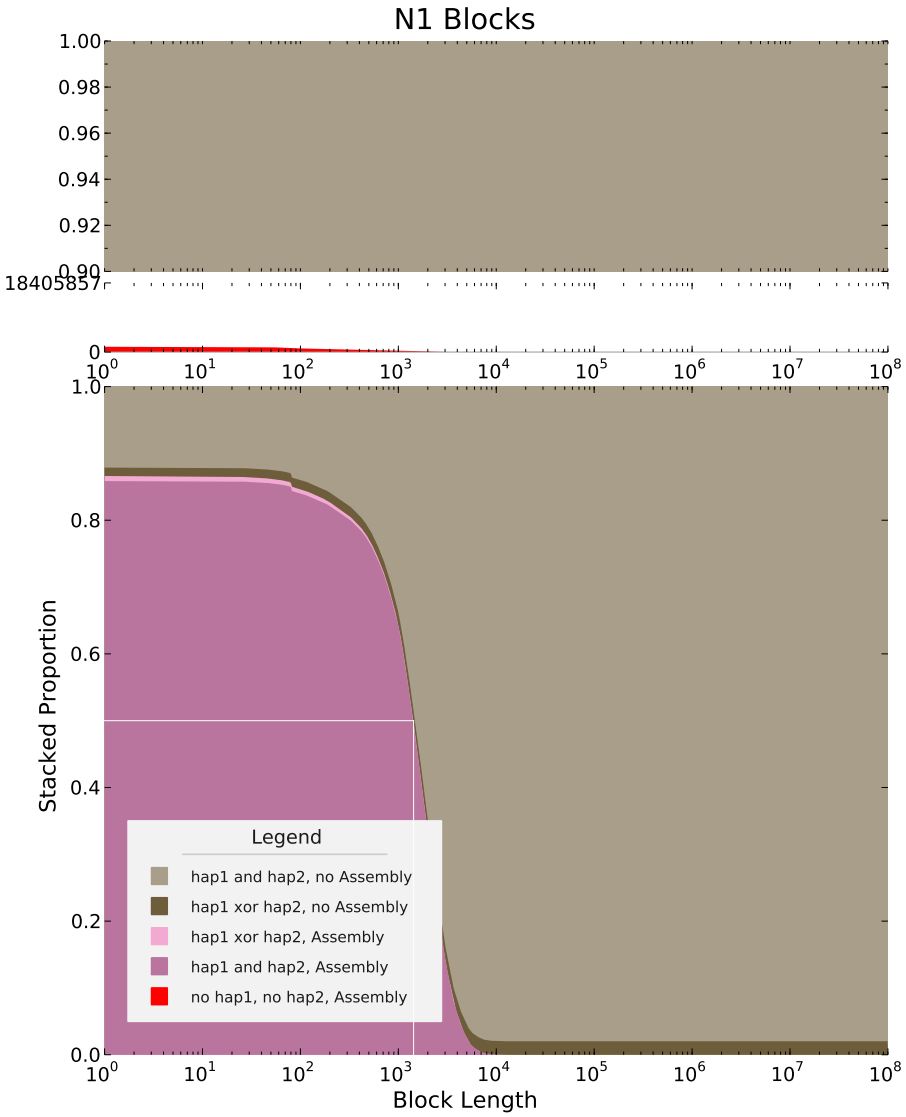


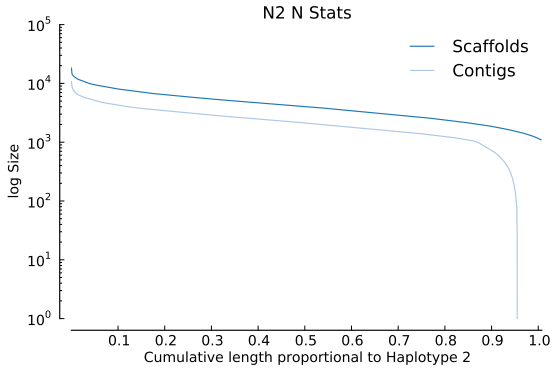
Figure 3.162: N1 blocks caption goes here.

N2

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
L1	0.83722	0.83727	0.83716	0.00000
N2	0.79226	0.79257	0.79195	0.00000
O1	0.78557	0.79017	0.78096	0.00558

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	33,787	1,082	1,884.00	2,811	3,356.00	4,288.00	18,206	1,989.20	113,389,083
Contigs	69,948	1	562.00	1,365	1,536.30	2,153.00	10,764	1,198.58	107,460,922

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	92,565,231 – 95,104,910	71,389,031 – 72,748,731	142,751,558.0 – 145,453,938.0	4,199 – 8,802
Heterozygous	356,453 – 386,257	262,951 – 274,546	525,602.0 – 548,516.0	14 – 48
Indel	1,708,456 – 2,080,270	590,343 – 693,819	1,178,962.0 – 1,384,798.0	844 – 1,110

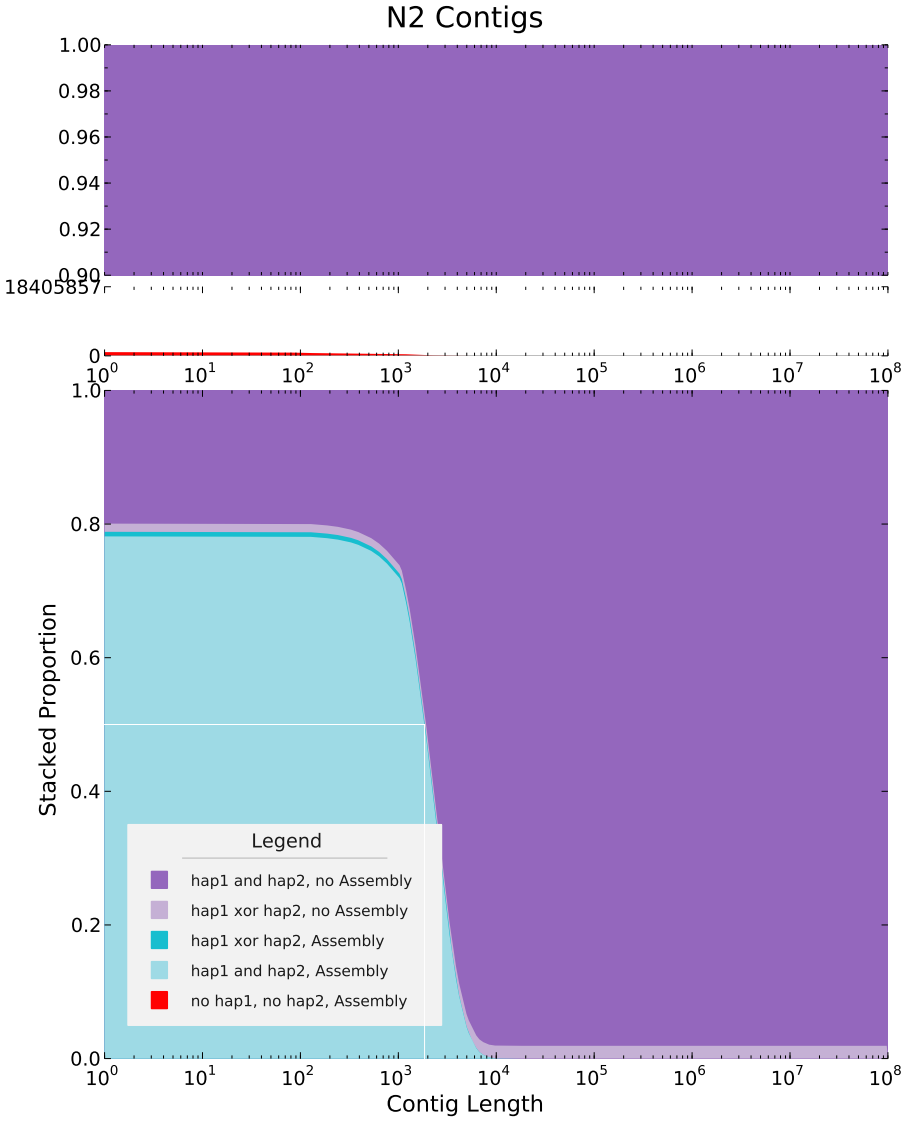


Figure 3.163: N2 contigs caption goes here.

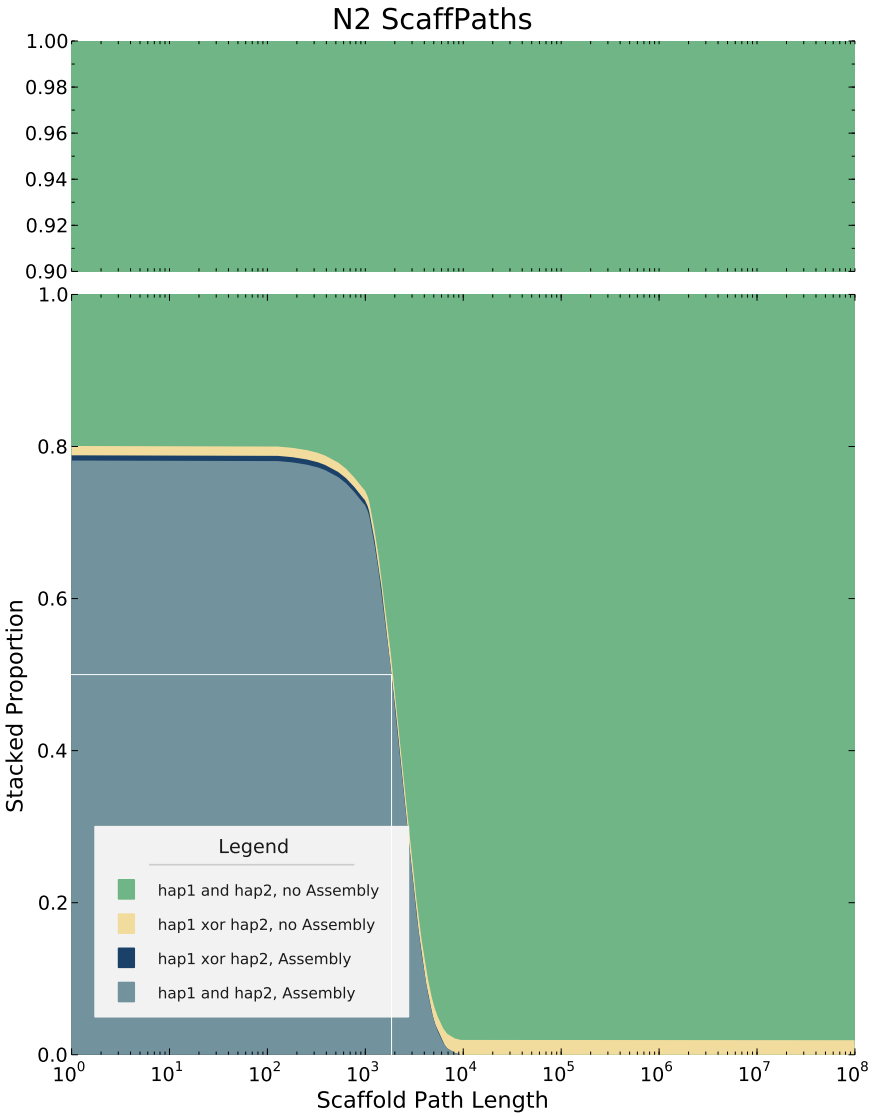


Figure 3.164: N2 scaffolds caption goes here.

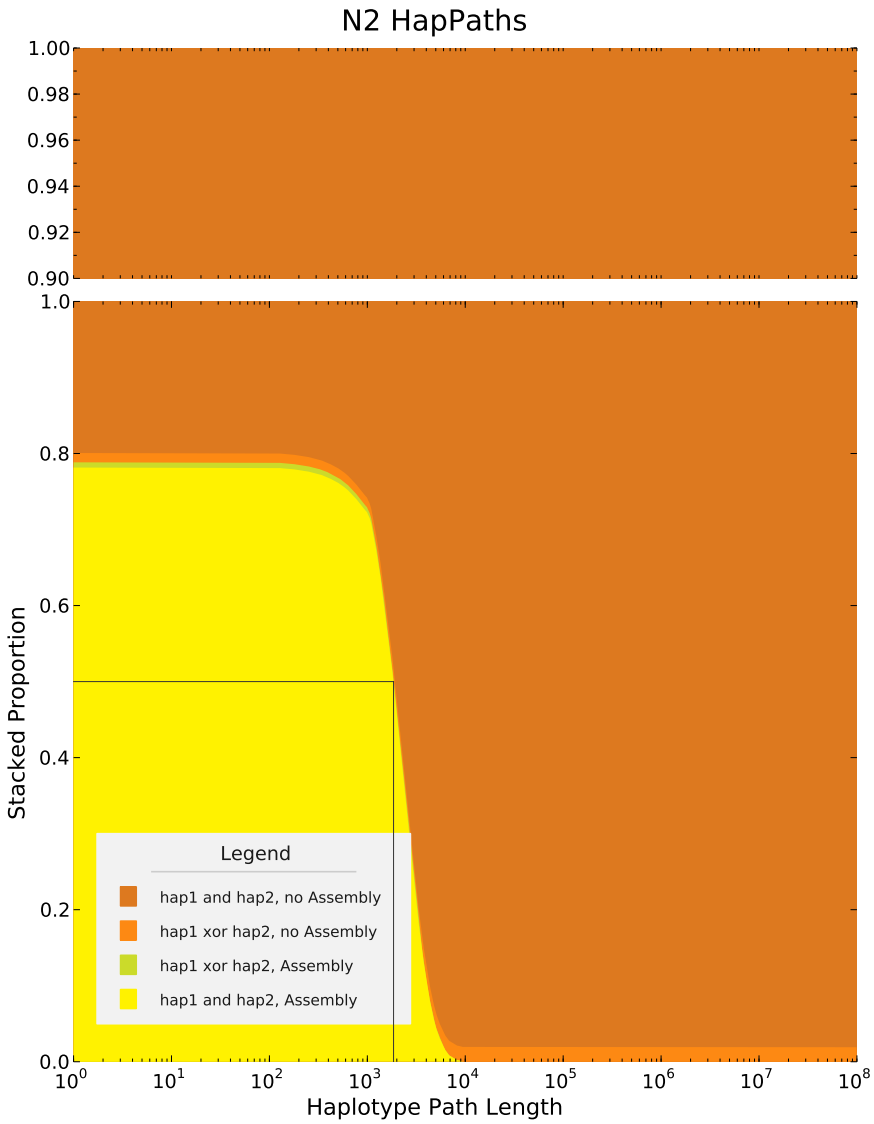


Figure 3.165: N2 hapPaths caption goes here.

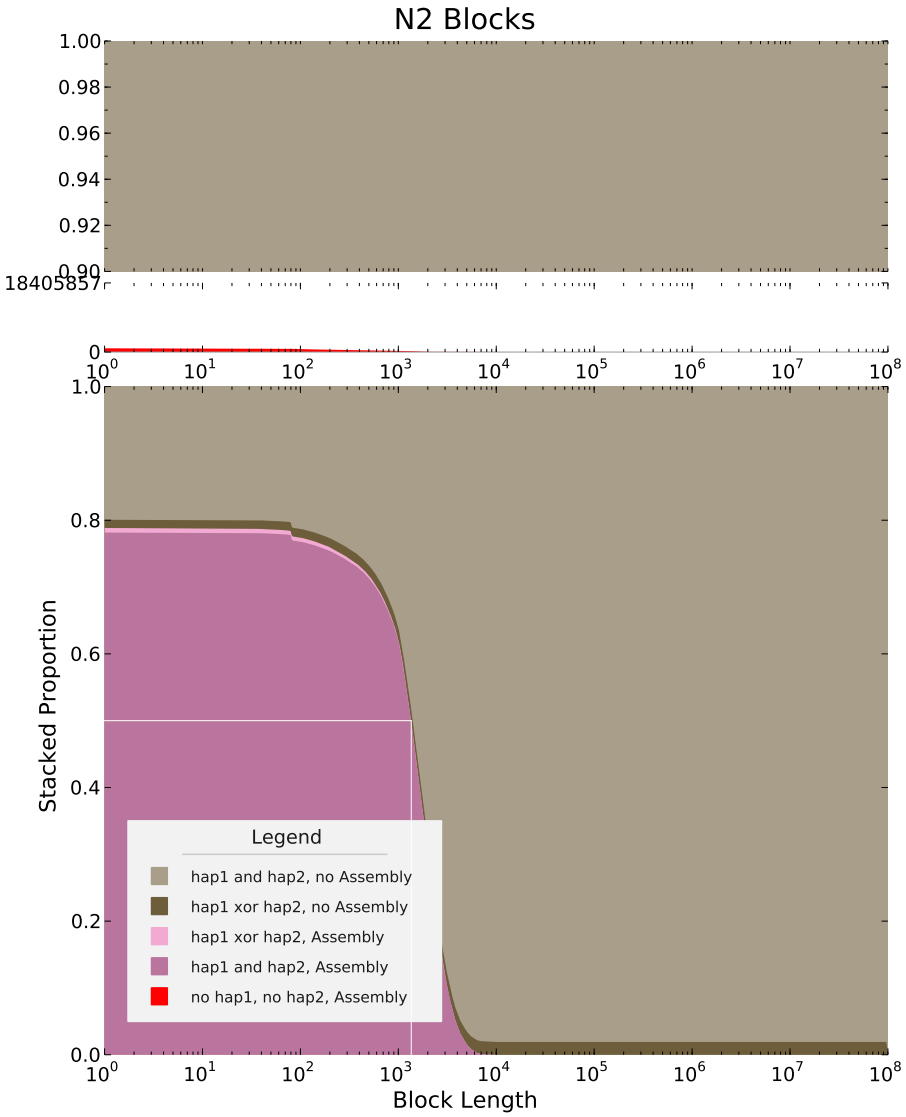


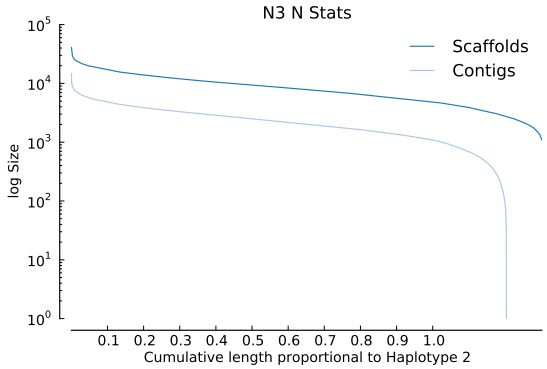
Figure 3.166: N2 blocks caption goes here.

N3

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
N1	0.87107	0.87121	0.87094	0.00000
N3	0.85657	0.85681	0.85634	0.00000
L1	0.83722	0.83727	0.83716	0.00000

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	24,298	1,082	2,881.00	5,016	6,029.43	7,998.75	40,835	4,102.13	146,503,047
Contigs	103,555	1	327.00	1,001	1,308.29	1,910.00	14,662	1,239.19	135,479,743

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	76,776,616 – 80,970,919	63,246,902 – 65,763,077	126,410,052.0 – 131,381,156.0	4,143 – 9,032
Heterozygous	297,319 – 331,460	234,954 – 250,468	468,150.0 – 497,438.0	20 – 49
Indel	1,679,463 – 2,052,370	530,923 – 653,535	1,059,928.0 – 1,301,104.0	770 – 1,120

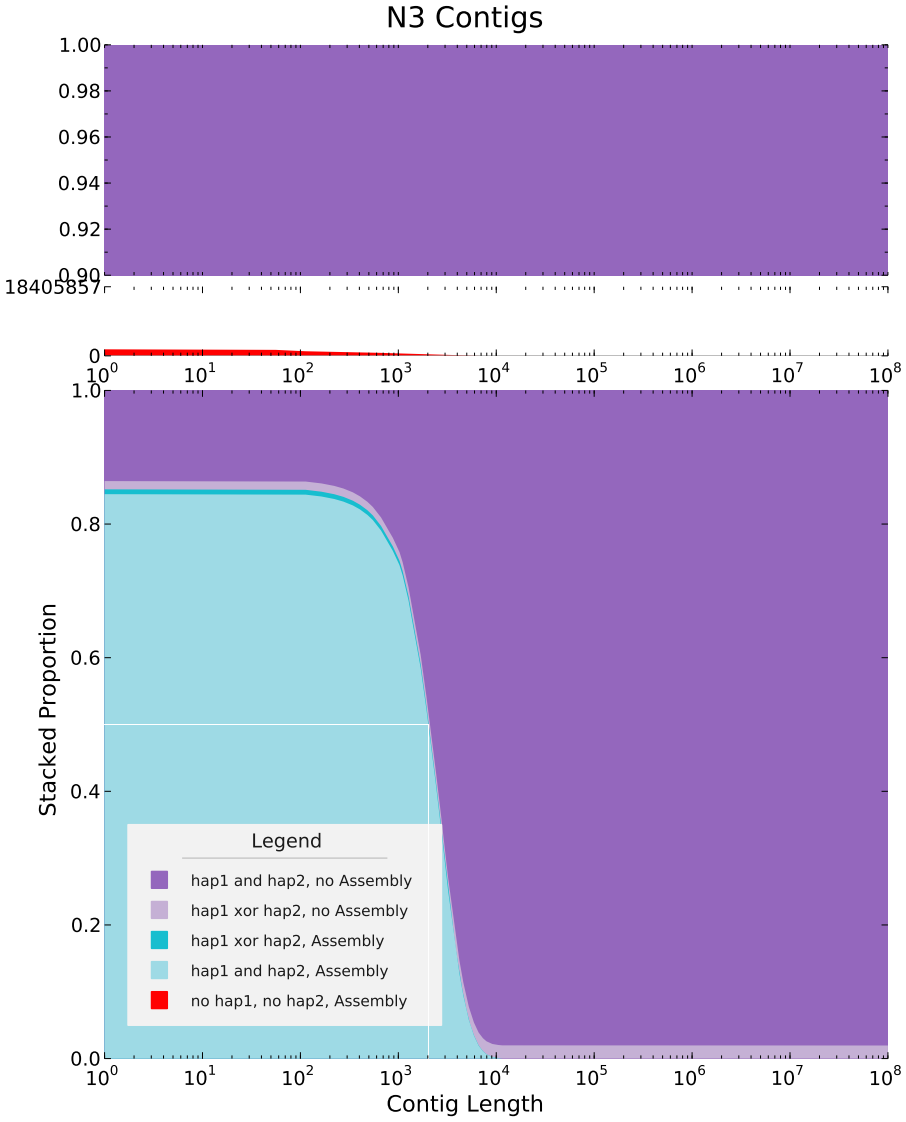


Figure 3.167: N3 contigs caption goes here.

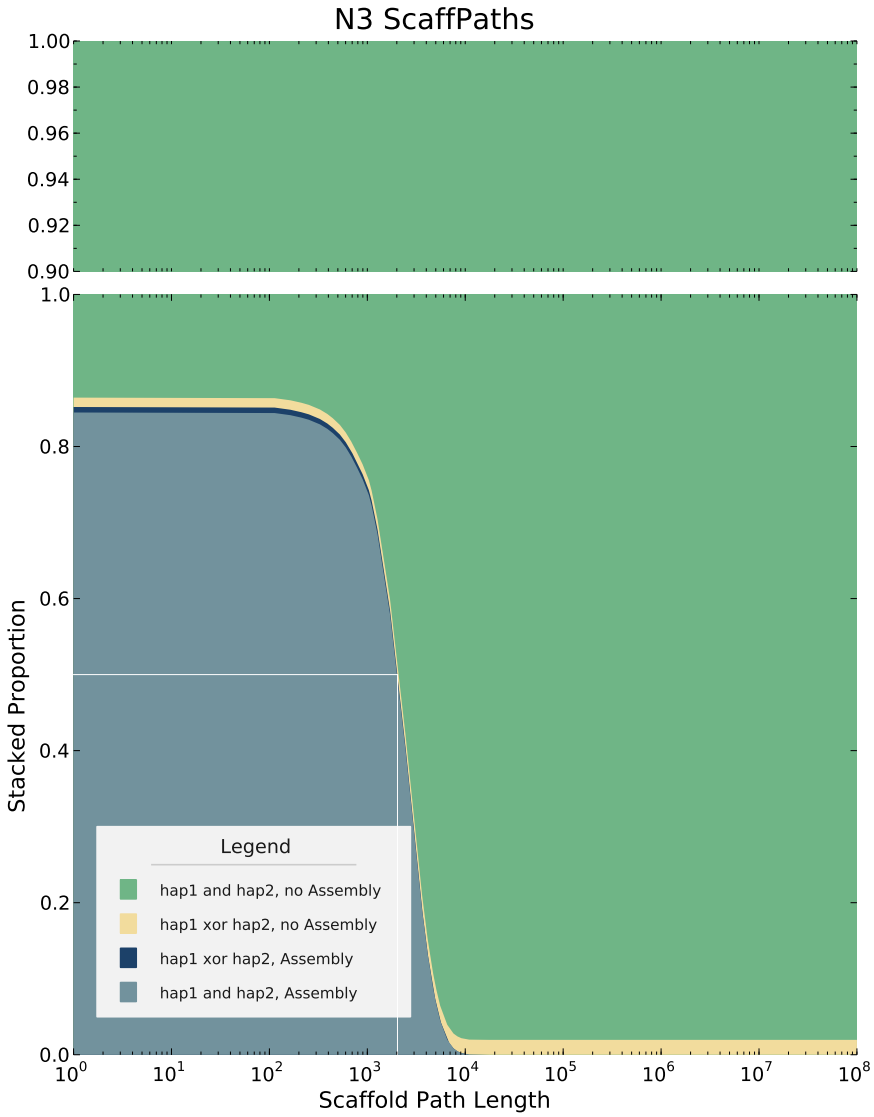


Figure 3.168: N3 scaffolds caption goes here.

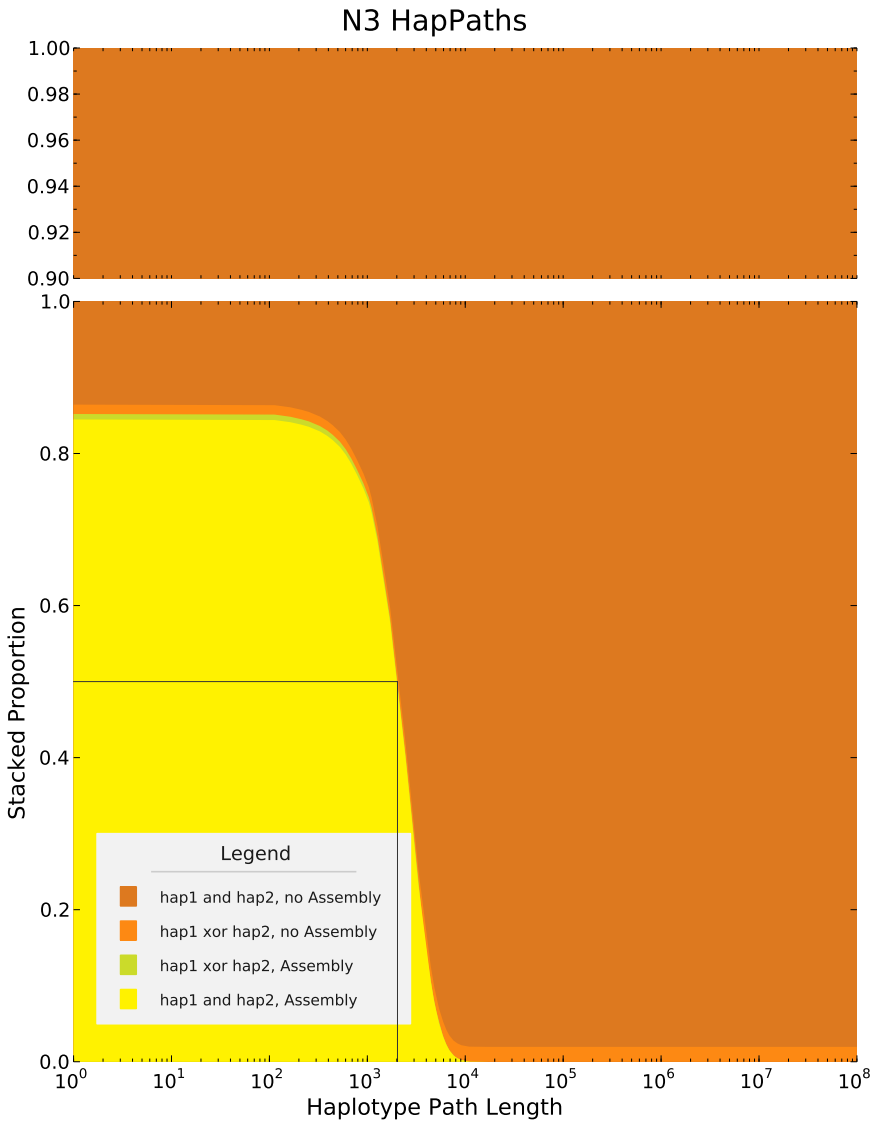


Figure 3.169: N3 hapPaths caption goes here.

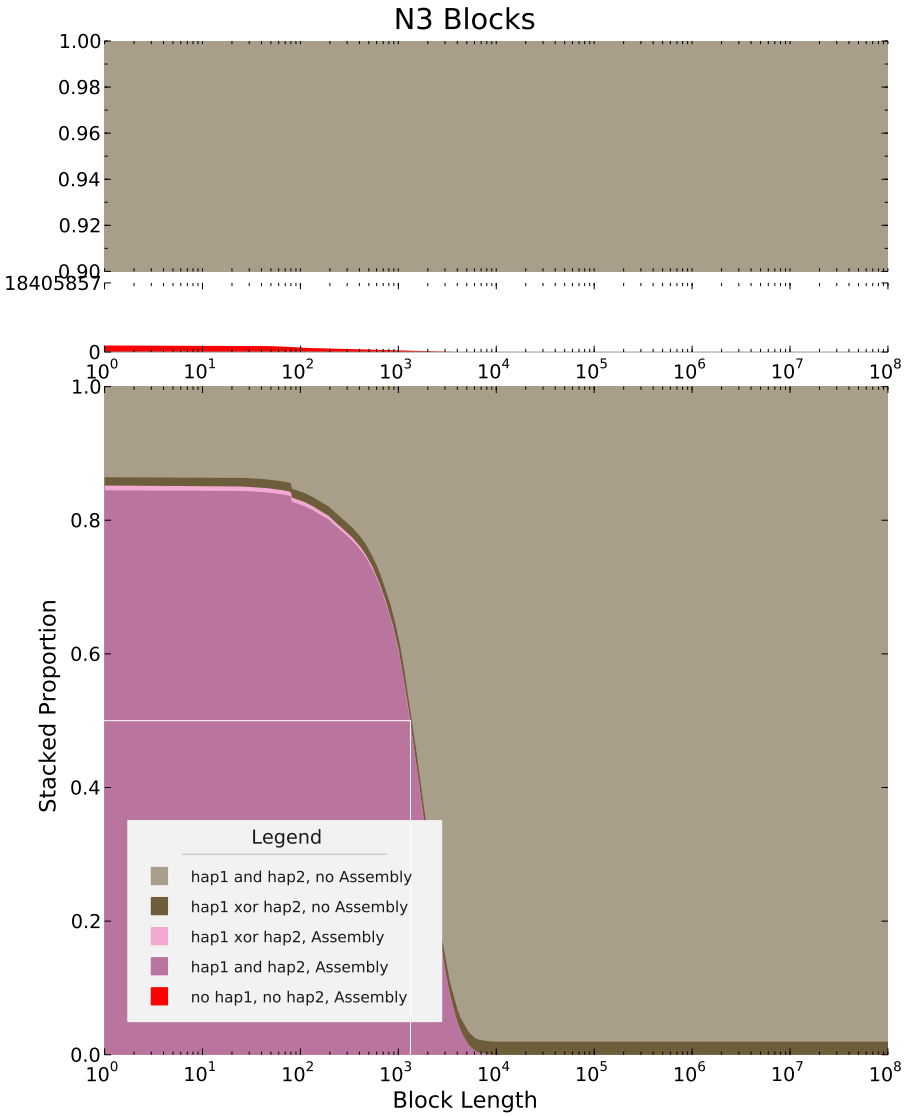


Figure 3.170: N3 blocks caption goes here.

3.2.15 O, KAS

Affiliation: Department of Computer Science, University of Chicago, USA
Contact: Fangfang Xia
Software: **Kiki**
Number of entries: 1

ID	Total	Hap 1	Hap 2	Bac
O1	0.78557	0.79017	0.78096	0.00558

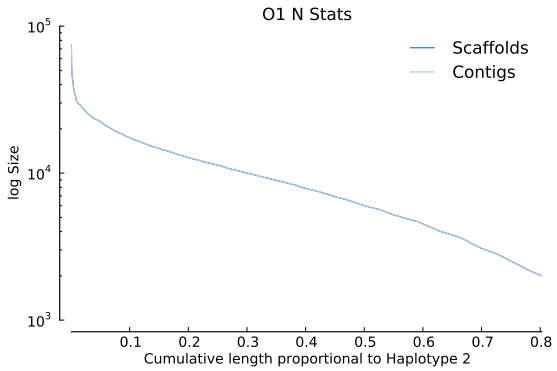
Assemblies:

O1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
L1	0.83722	0.83727	0.83716	0.00000
N2	0.79226	0.79257	0.79195	0.00000
O1	0.78557	0.79017	0.78096	0.00558

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	14,994	2,000	2,839.00	4,360	6,024.66	7,605.00	74,215	4,690.22	90,333,710
Contigs	14,994	2,000	2,839.00	4,360	6,024.66	7,605.00	74,215	4,690.22	90,333,710

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,045,903 – 108,951,478	85,158,225 – 85,811,948	170,293,418.0 – 171,572,566.0	11,516 – 25,665
Heterozygous	425,297 – 434,585	333,807 – 341,427	667,550.0 – 682,724.0	32 – 65
Indel	2,405,864 – 2,797,300	1,038,036 – 1,202,358	2,073,996.0 – 2,398,708.0	1,038 – 3,004

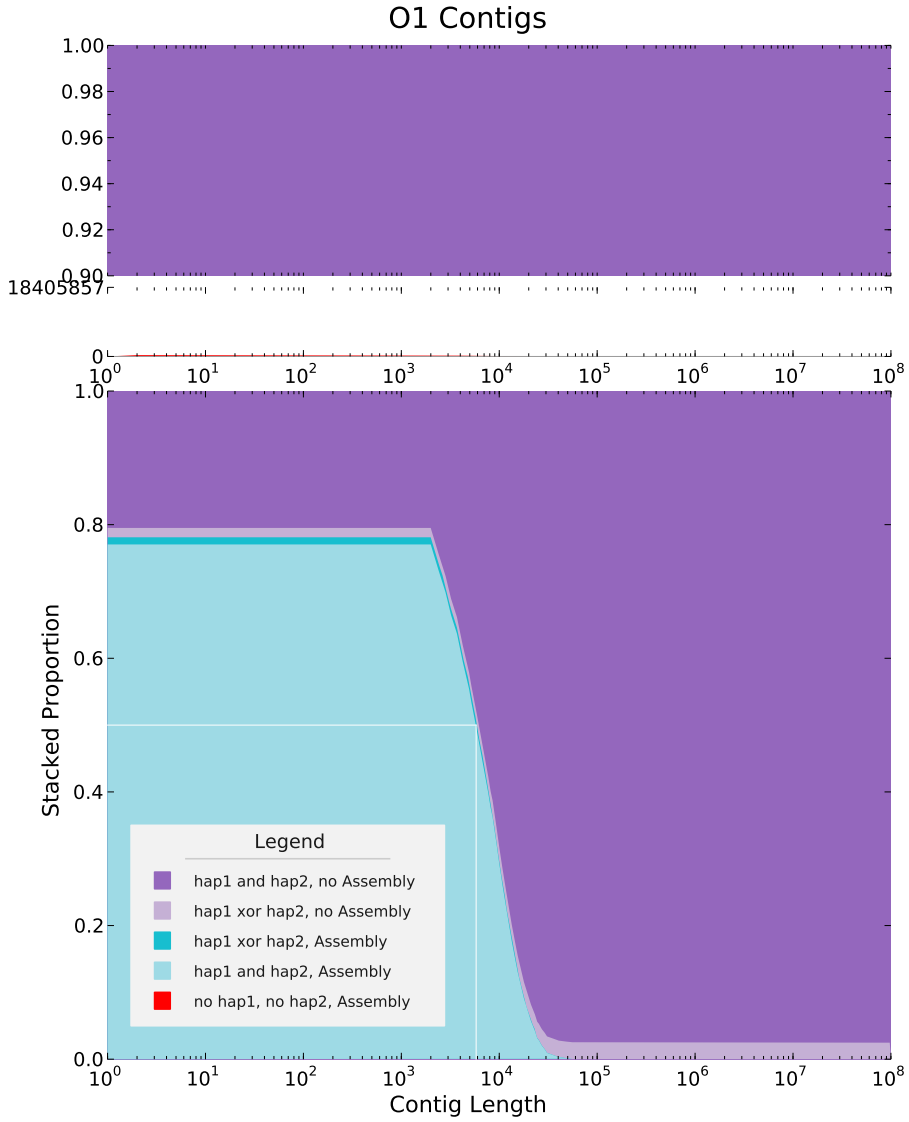


Figure 3.171: O1 contigs caption goes here.

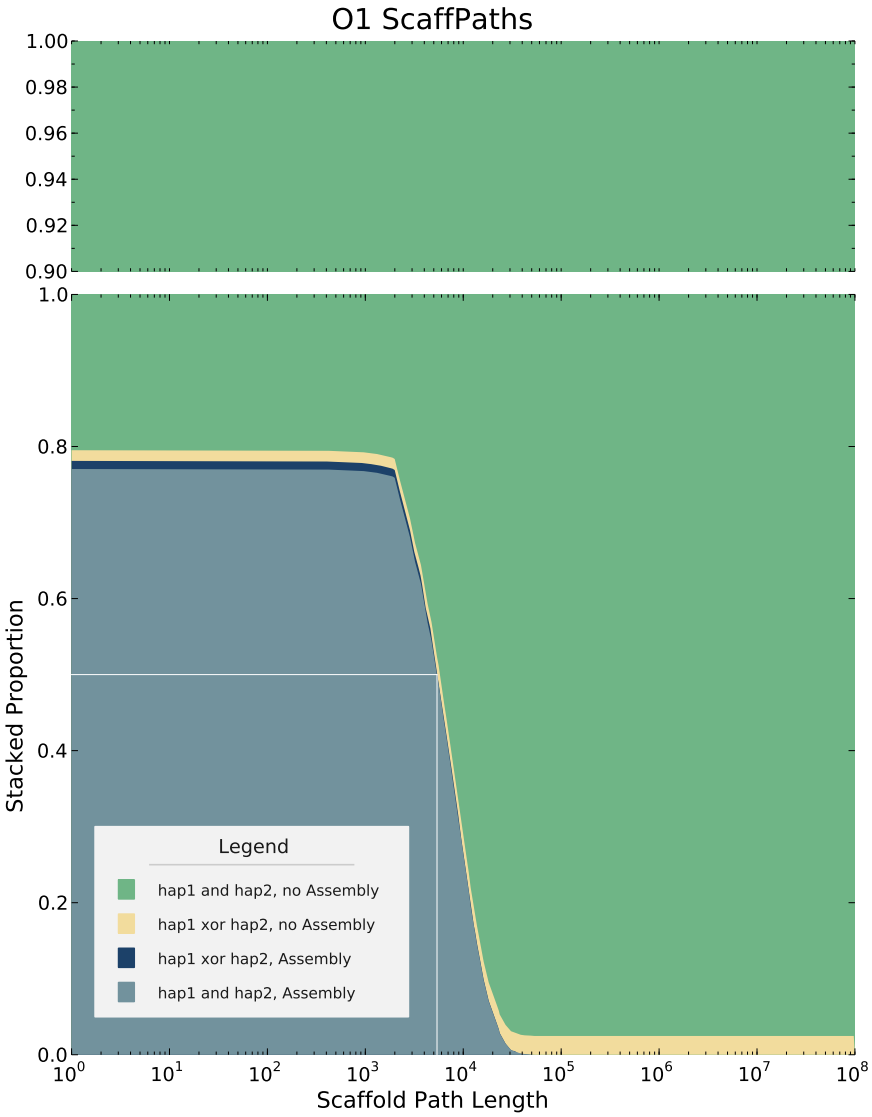


Figure 3.172: O1 scaffolds caption goes here.

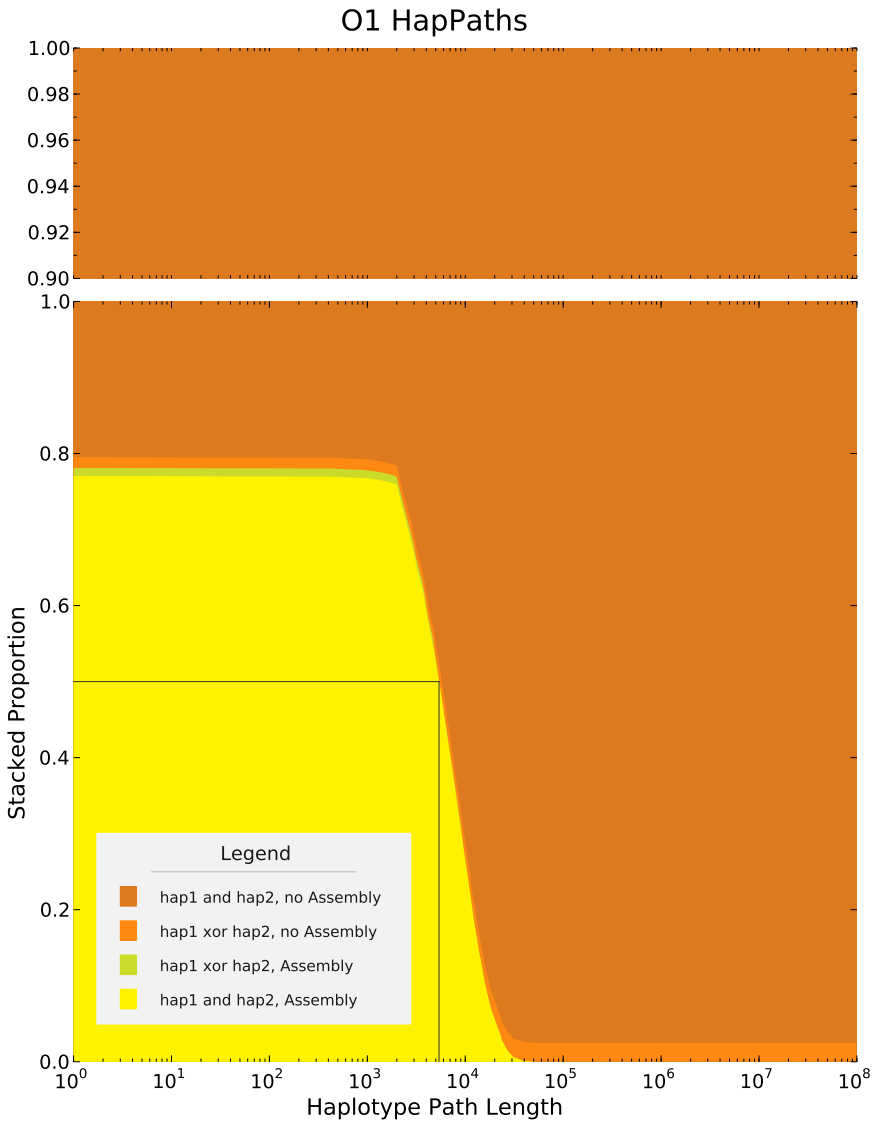


Figure 3.173: O1 hapPaths caption goes here.

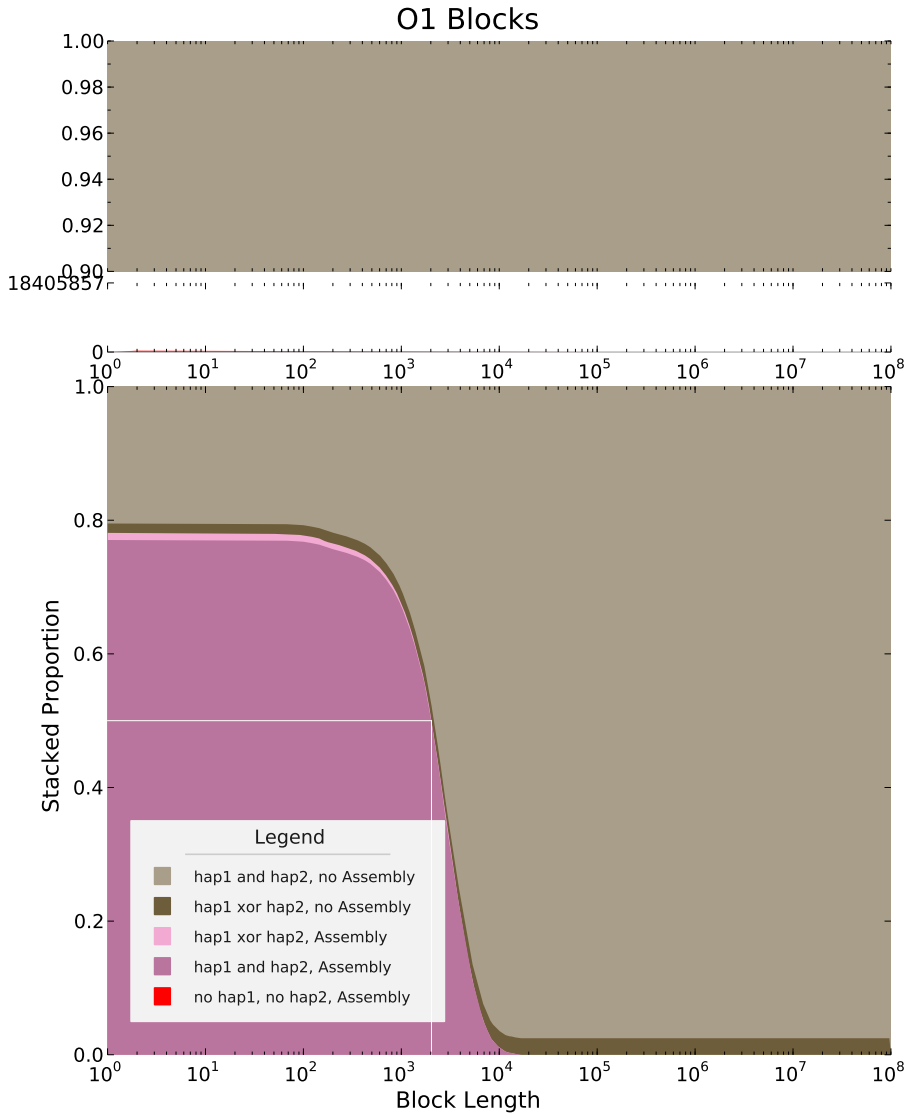


Figure 3.174: O1 blocks caption goes here.

3.2.16 P, BGI-Shenzhen

Affiliation: BGI, China
Contact: Zhenyu Li
Software: **SOAPdenovo**
Number of entries: 1

ID	Total	Hap 1	Hap 2	Bac
P1	0.98852	0.98881	0.98823	0.00000

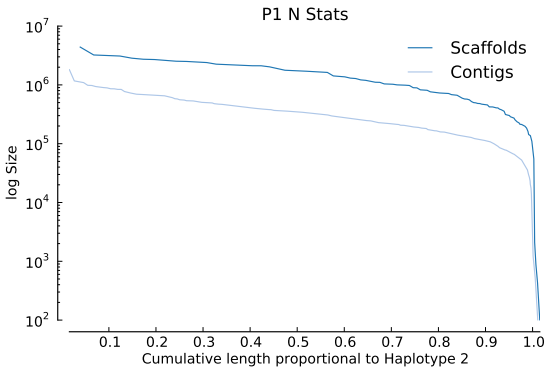
Assemblies:

P1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
P1	0.98852	0.98881	0.98823	0.00000
B1	0.98694	0.98719	0.98668	0.99790
F5	0.98691	0.98727	0.98653	0.99934

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	4,143	100	116.00	175	27,578.70	431.00	4,397,505	217,050.00	114,258,553
Contigs	4,566	100	118.00	205	24,919.20	609.75	1,814,562	99,098.96	113,781,081

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	107,821,484 – 108,416,677	107,475,470 – 108,037,854	214,947,712.0 – 216,069,946.0	1,500 – 2,532
Heterozygous	421,708 – 432,050	419,968 – 429,848	839,904.0 – 859,632.0	11 – 20
Indel	3,155,268 – 3,540,700	1,521,634 – 1,716,959	3,038,006.0 – 3,426,412.0	2,630 – 3,620

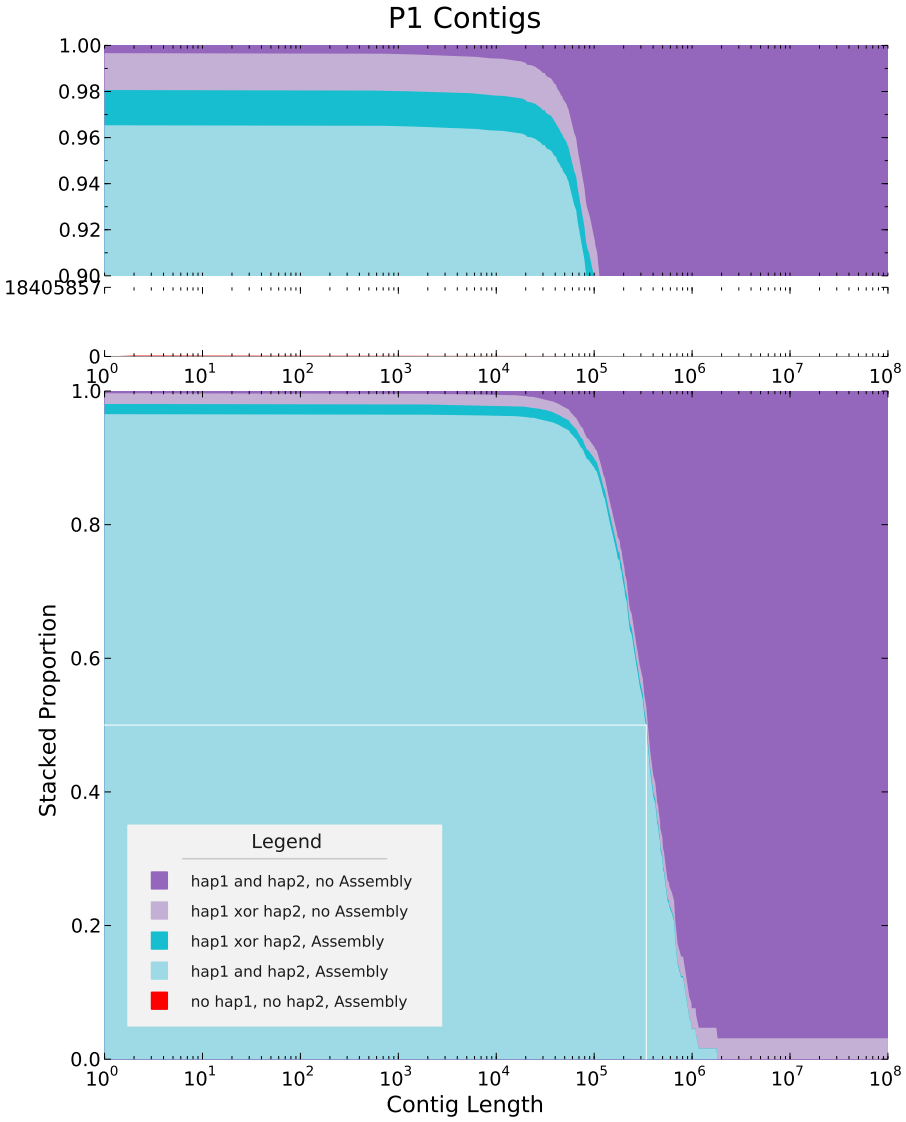


Figure 3.175: P1 contigs caption goes here.

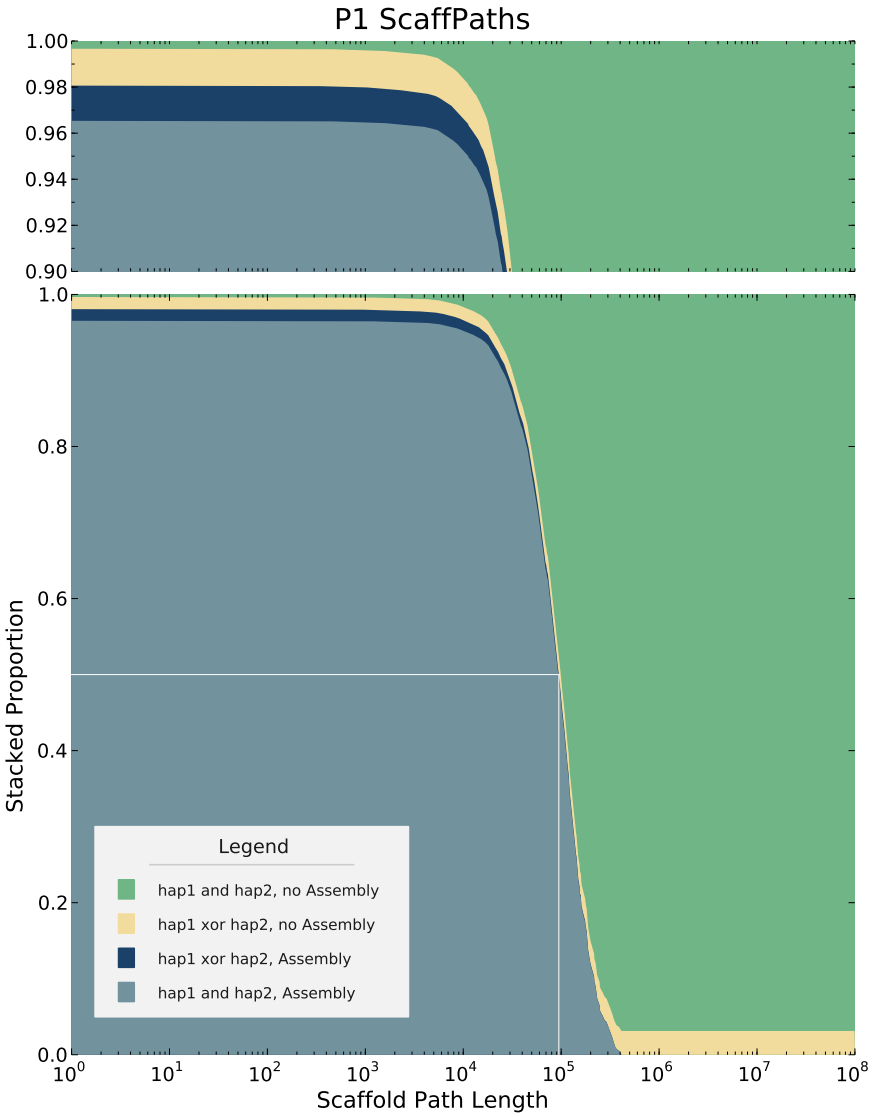


Figure 3.176: P1 scaffolds caption goes here.

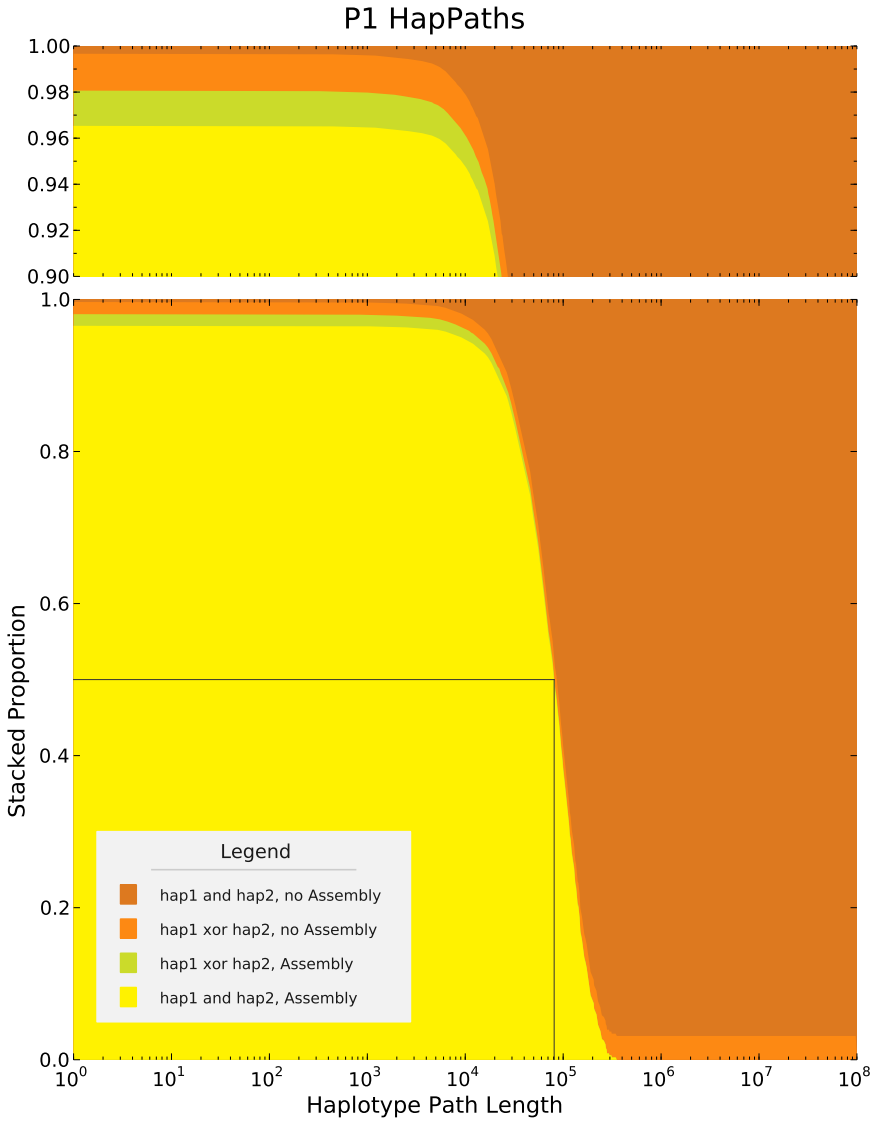


Figure 3.177: P1 hapPaths caption goes here.

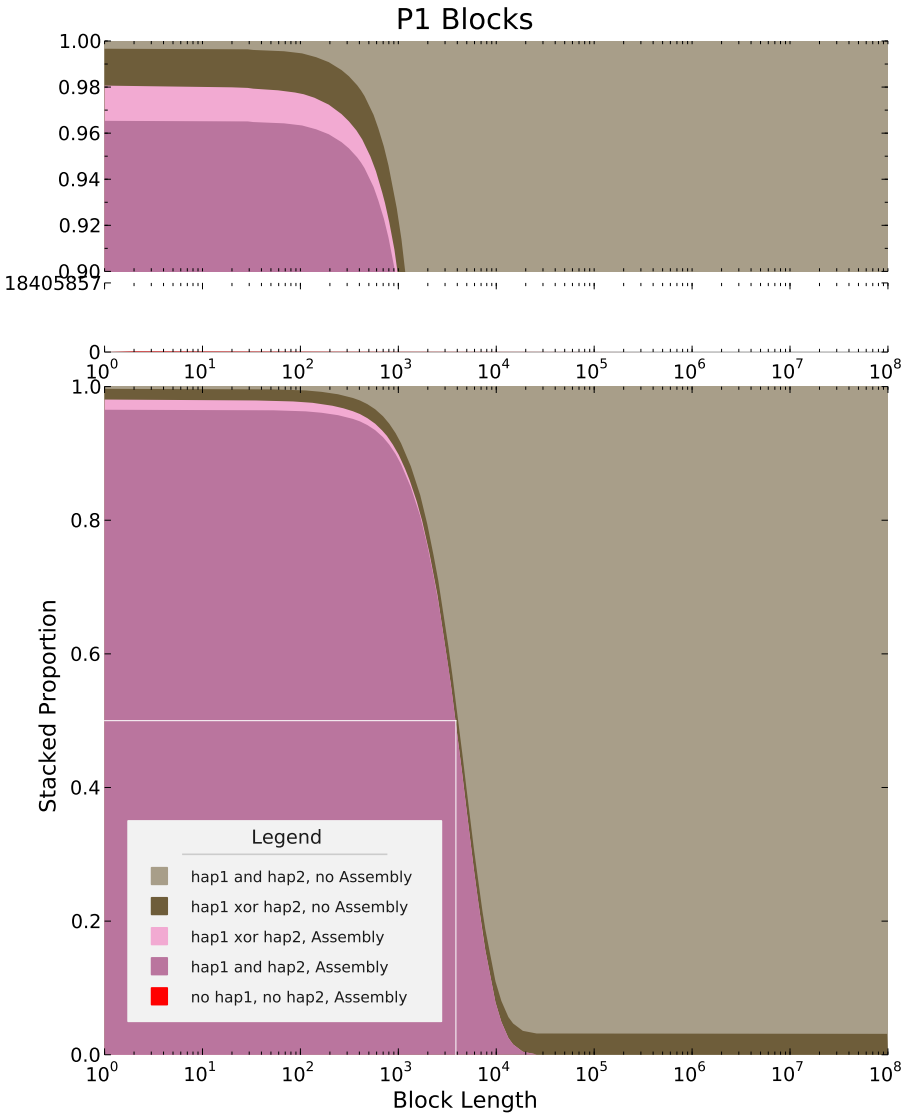


Figure 3.178: P1 blocks caption goes here.

3.2.17 Q, ALLPATHS Assembly Team

Affiliation: Broad Institute
Contact: David Jaffe
Software: **ALLPATHS-LG**
Number of entries: 1

ID	Total	Hap 1	Hap 2	Bac
Q1	0.98325	0.98337	0.98311	0.68864

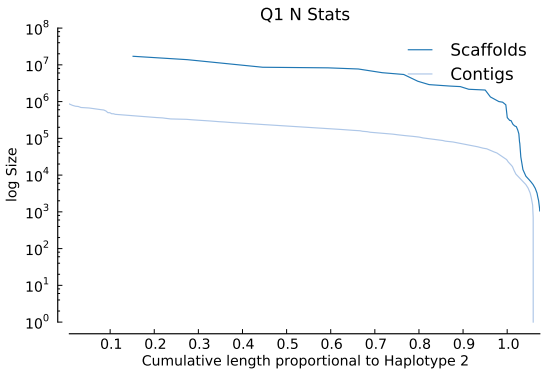
Assemblies:

Q1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
I2	0.98467	0.98511	0.98424	0.99857
Q1	0.98325	0.98337	0.98311	0.68864
K1	0.98306	0.98309	0.98302	0.11258

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	990	1,004	2,328.25	4,511	122,076.64	7,338.50	17,101,185	997,651.65	120,855,878
Contigs	1,946	1	3,337.75	8,551	61,240.55	75,652.50	858,807	105,851.53	119,174,112

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	105,634,099 – 106,191,597	104,400,809 – 104,915,082	208,796,220.0 – 209,820,946.0	386 – 815
Heterozygous	415,374 – 425,609	409,789 – 418,816	160,952.0 – 173,924.0	23 – 27
Indel	2,060,212 – 2,423,335	937,734 – 1,245,544	1,871,056.0 – 2,483,110.0	426 – 1,186

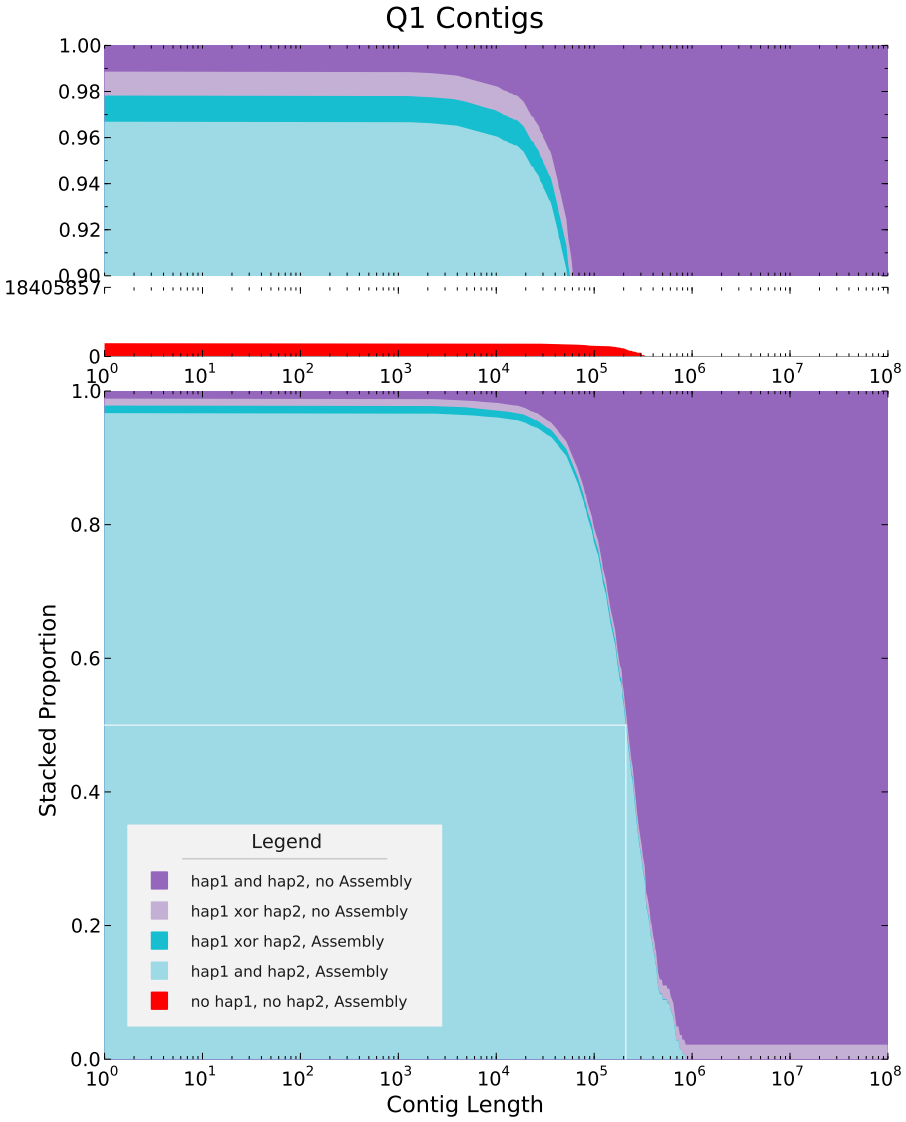


Figure 3.179: Q1 contigs caption goes here.

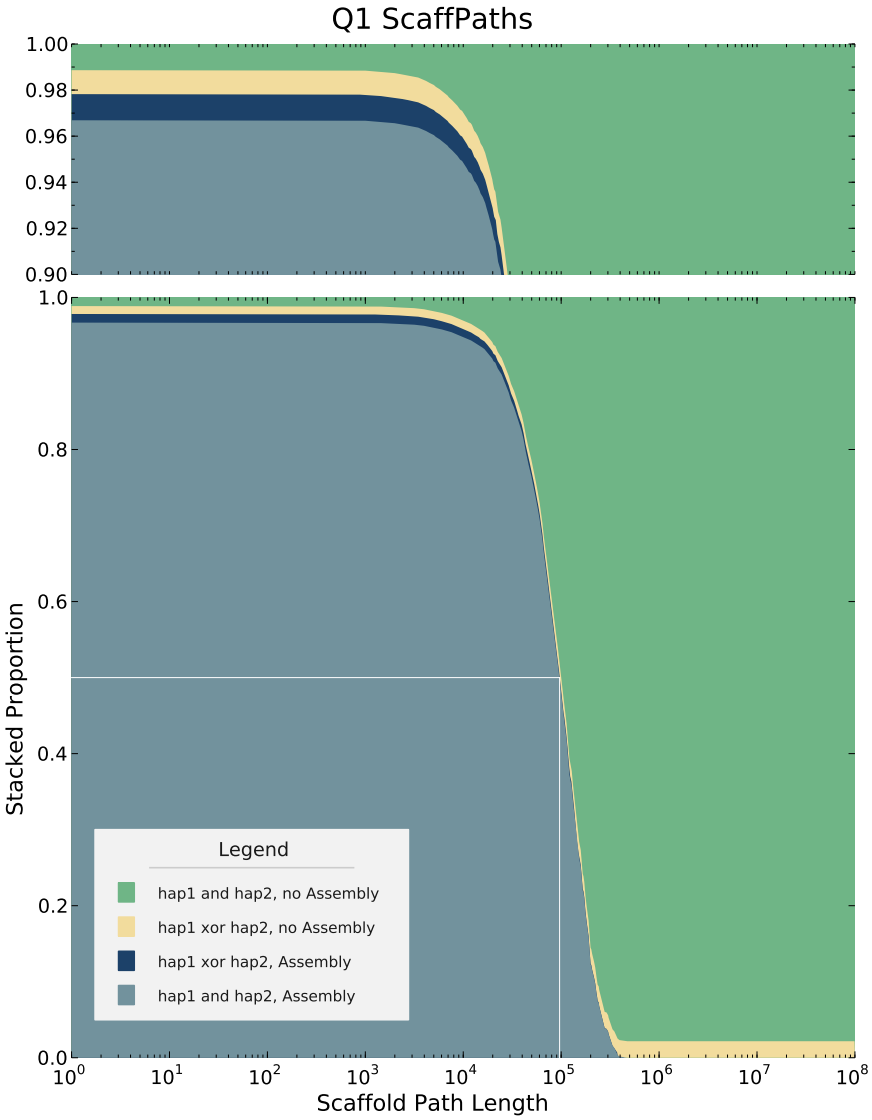


Figure 3.180: Q1 scaffolds caption goes here.

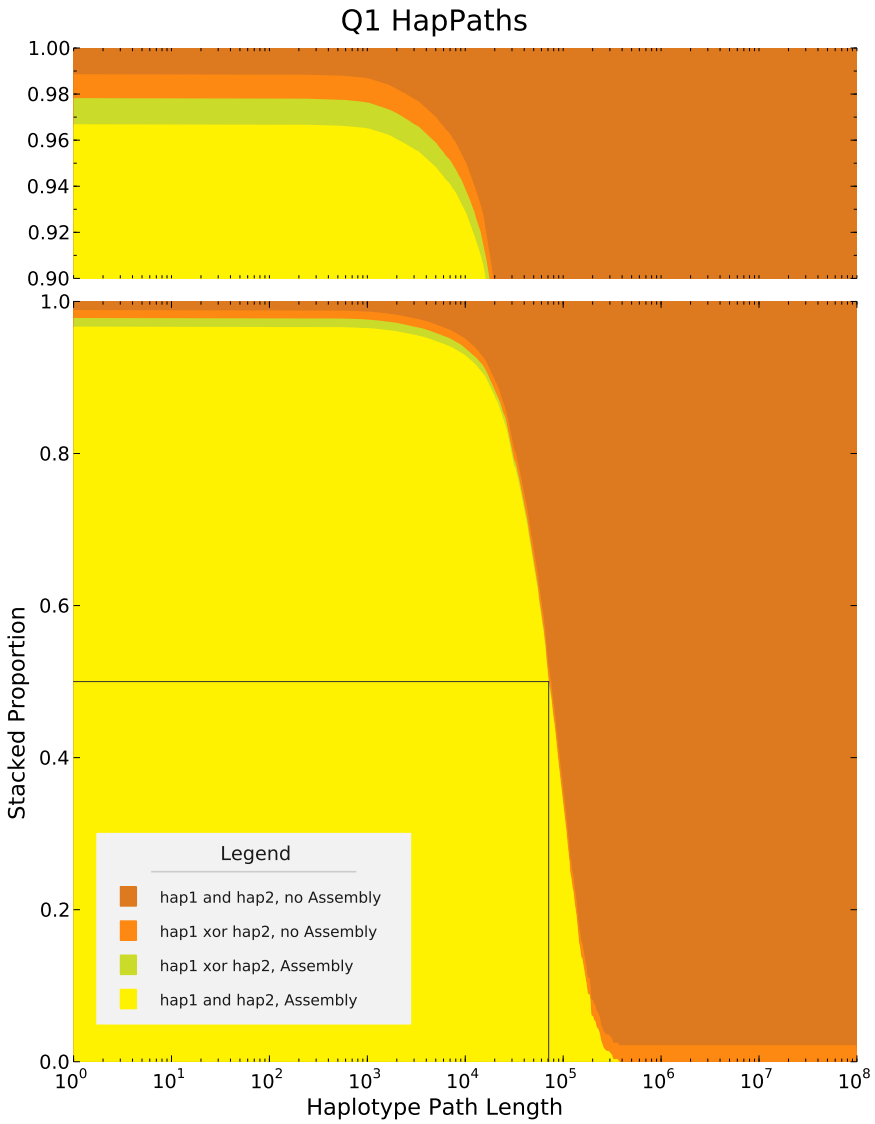


Figure 3.181: Q1 hapPaths caption goes here.

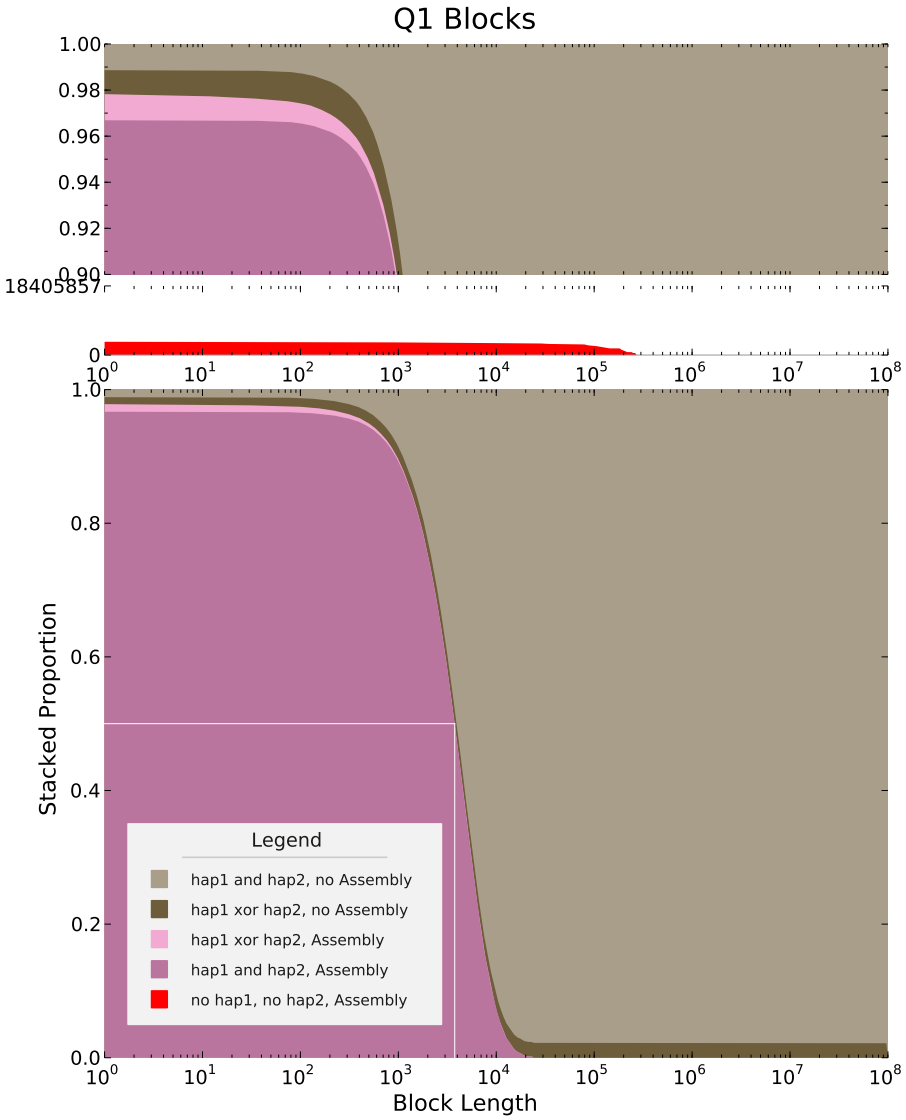


Figure 3.182: Q1 blocks caption goes here.

3.2.18 V, Auto

Affiliation: Auto
Contact: Auto
Software: **Velvet**
Number of entries: 6

ID	Total	Hap 1	Hap 2	Bac
V5	0.96373	0.96390	0.96357	0.99559
V6	0.96157	0.96183	0.96132	0.99560
V4	0.96153	0.96180	0.96128	0.99789

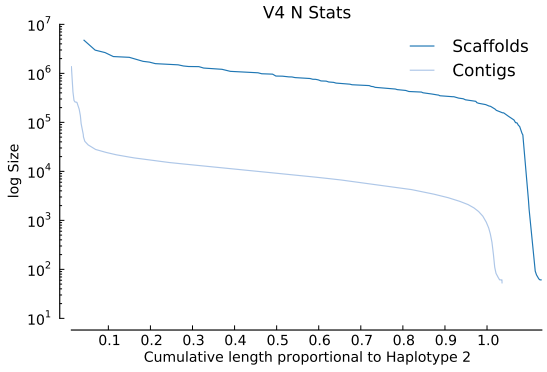
Assemblies:

V4

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
V6	0.96157	0.96183	0.96132	0.99560
V4	0.96153	0.96180	0.96128	0.99789
E1	0.96089	0.96125	0.96053	0.99725

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	32,617	61	61.00	69	3,899.37	85.00	4,743,985	64,313.71	127,185,684
Contigs	51,760	51	64.00	92	2,251.73	2,565.00	1,368,779	8,158.55	116,549,762

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,757,572 – 110,363,899	105,765,024 – 106,658,751	211,502,070.0 – 213,213,864.0	7,466 – 12,342
Heterozygous	422,165 – 440,464	407,856 – 417,220	814,982.0 – 831,926.0	20 – 29
Indel	1,551,498 – 1,938,567	621,878 – 789,930	1,241,562.0 – 1,568,298.0	1,036 – 1,565

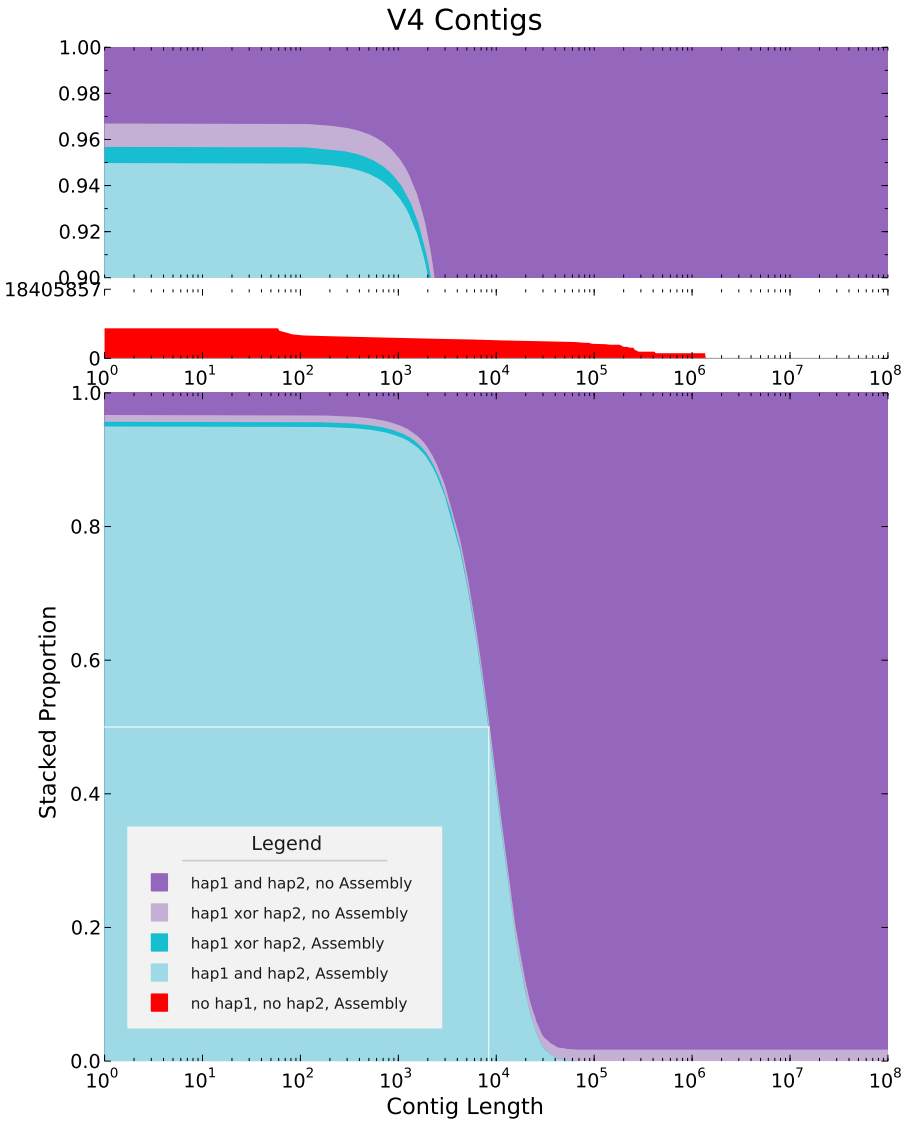


Figure 3.183: V4 contigs caption goes here.

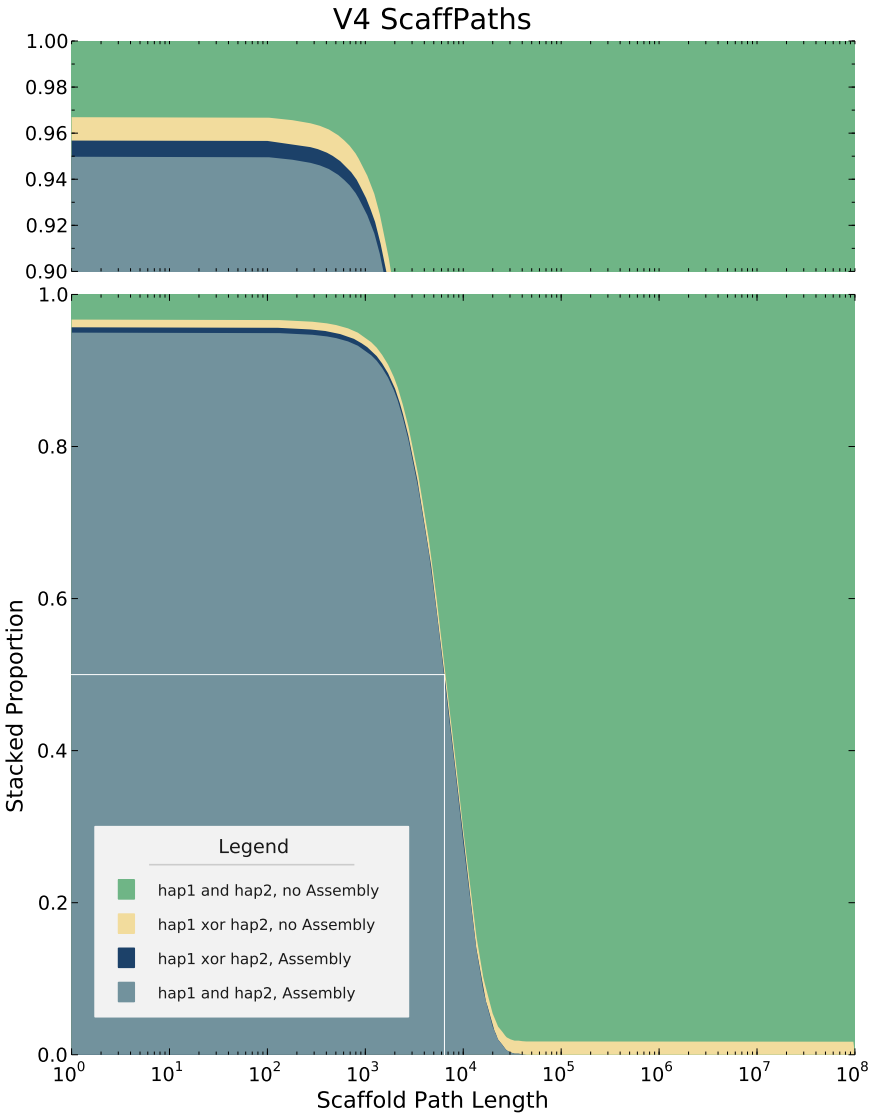


Figure 3.184: V4 scaffolds caption goes here.

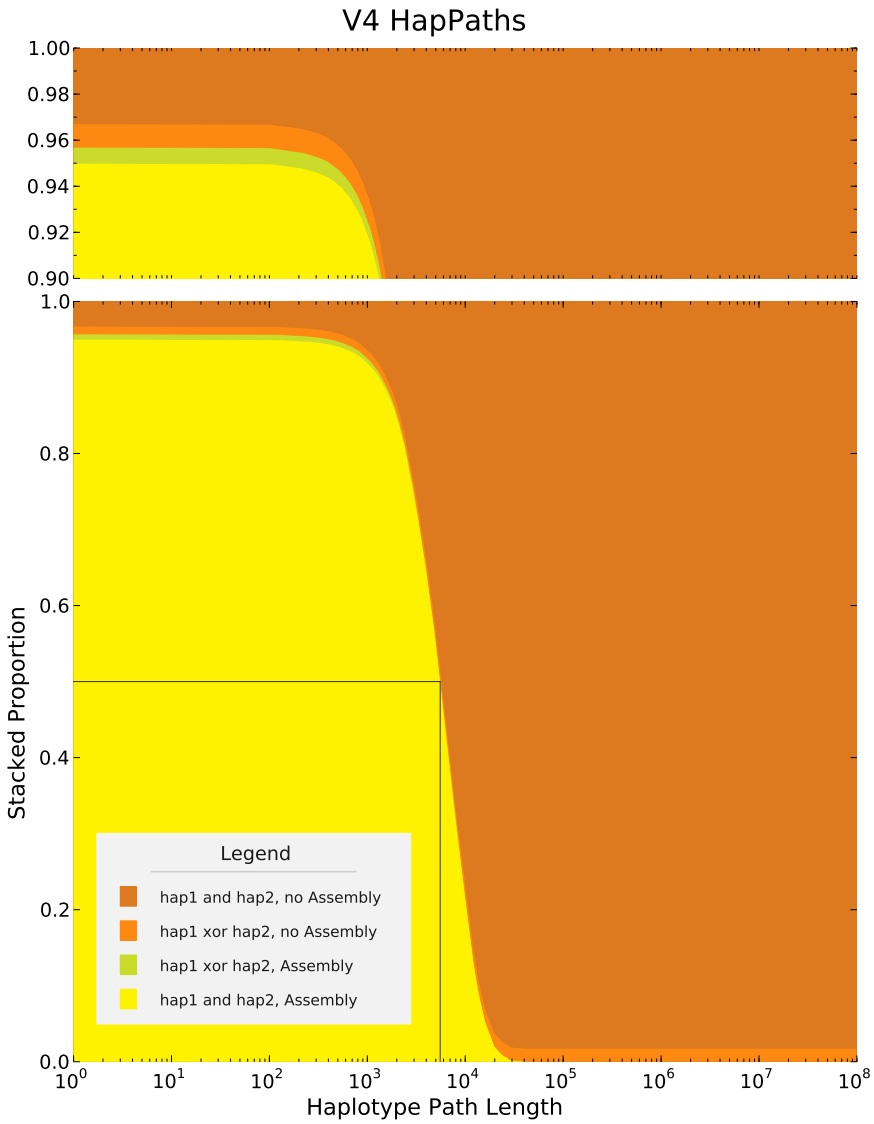


Figure 3.185: V4 hapPaths caption goes here.

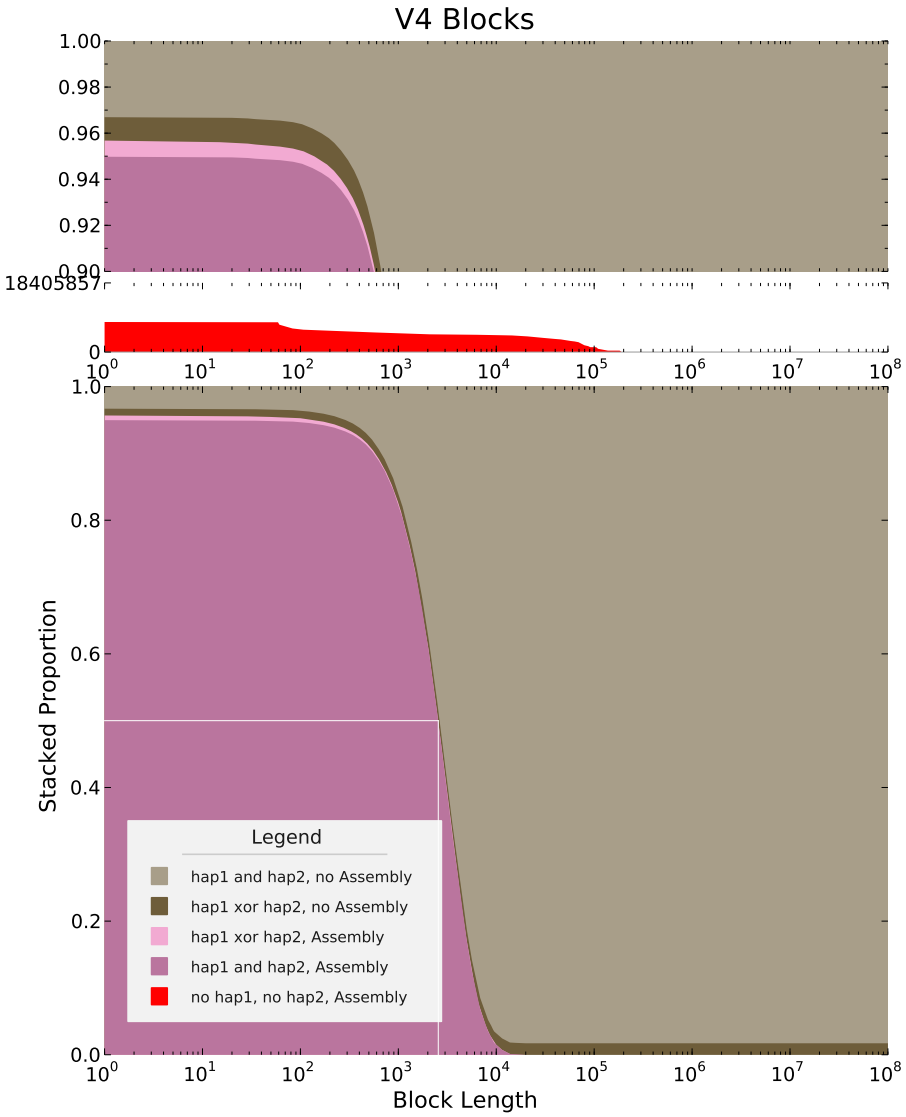


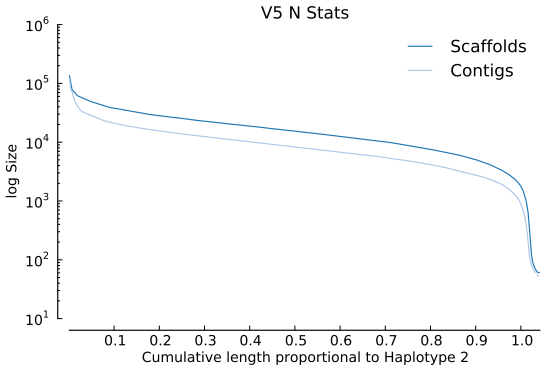
Figure 3.186: V4 blocks caption goes here.

V5

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
W3	0.96751	0.96771	0.96728	0.99810
V5	0.96373	0.96390	0.96357	0.99559
V6	0.96157	0.96183	0.96132	0.99560

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	44,364	61	62.00	78	2,642.68	1,325.25	134,380	6,489.21	117,239,847
Contigs	53,949	51	65.00	98	2,163.28	2,610.00	121,503	4,304.71	116,706,574

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,872,432 – 110,364,465	106,062,526 – 106,930,269	212,103,782.0 – 213,791,104.0	7,085 – 11,086
Heterozygous	422,036 – 440,269	408,773 – 418,162	817,274.0 – 835,204.0	16 – 25
Indel	1,607,768 – 1,999,314	657,490 – 833,806	1,312,662.0 – 1,658,304.0	1,110 – 1,655

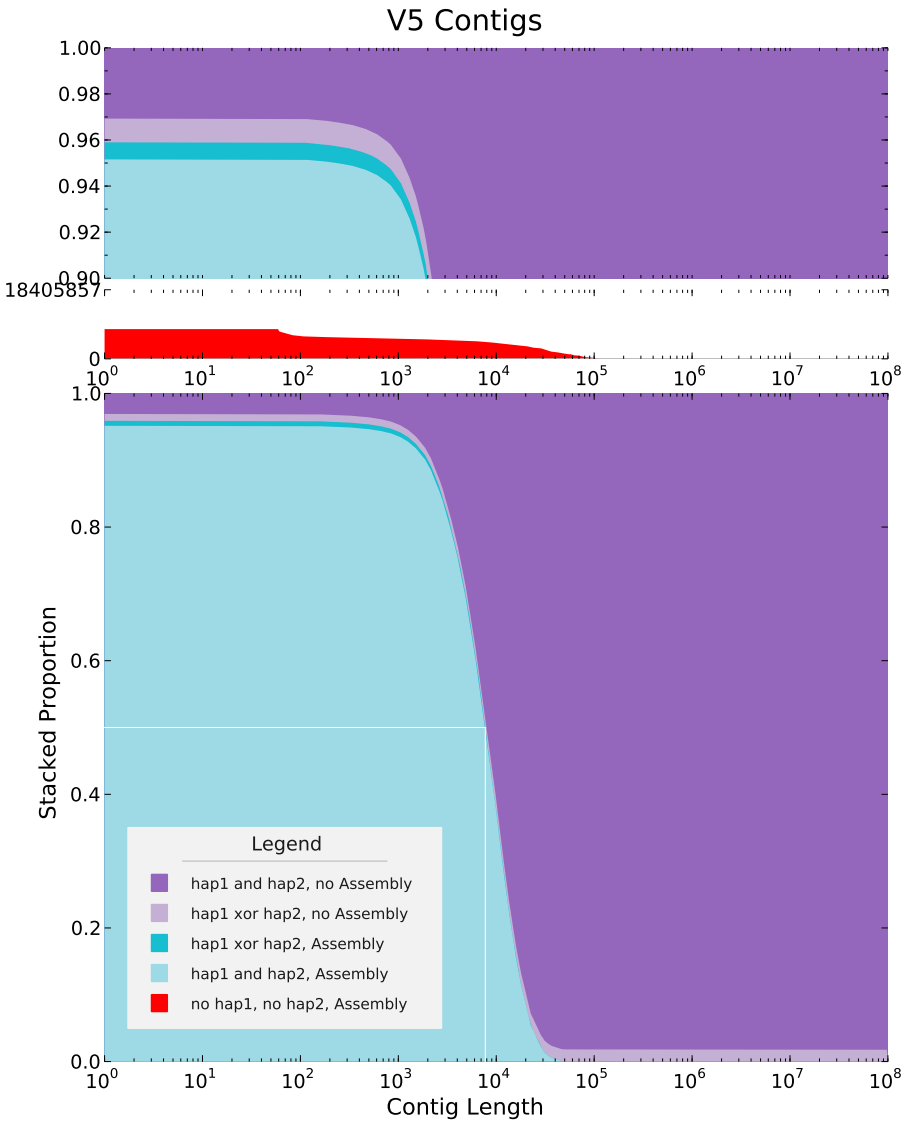


Figure 3.187: V5 contigs caption goes here.

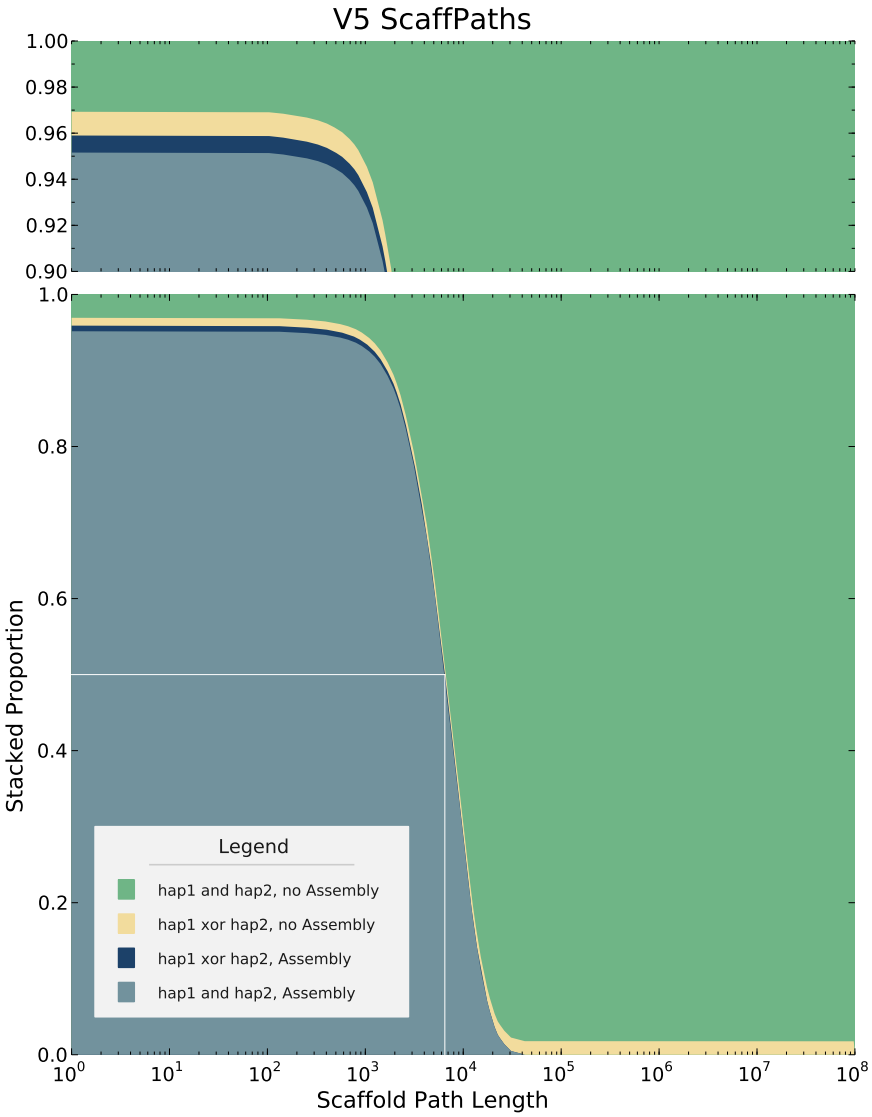


Figure 3.188: V5 scaffolds caption goes here.

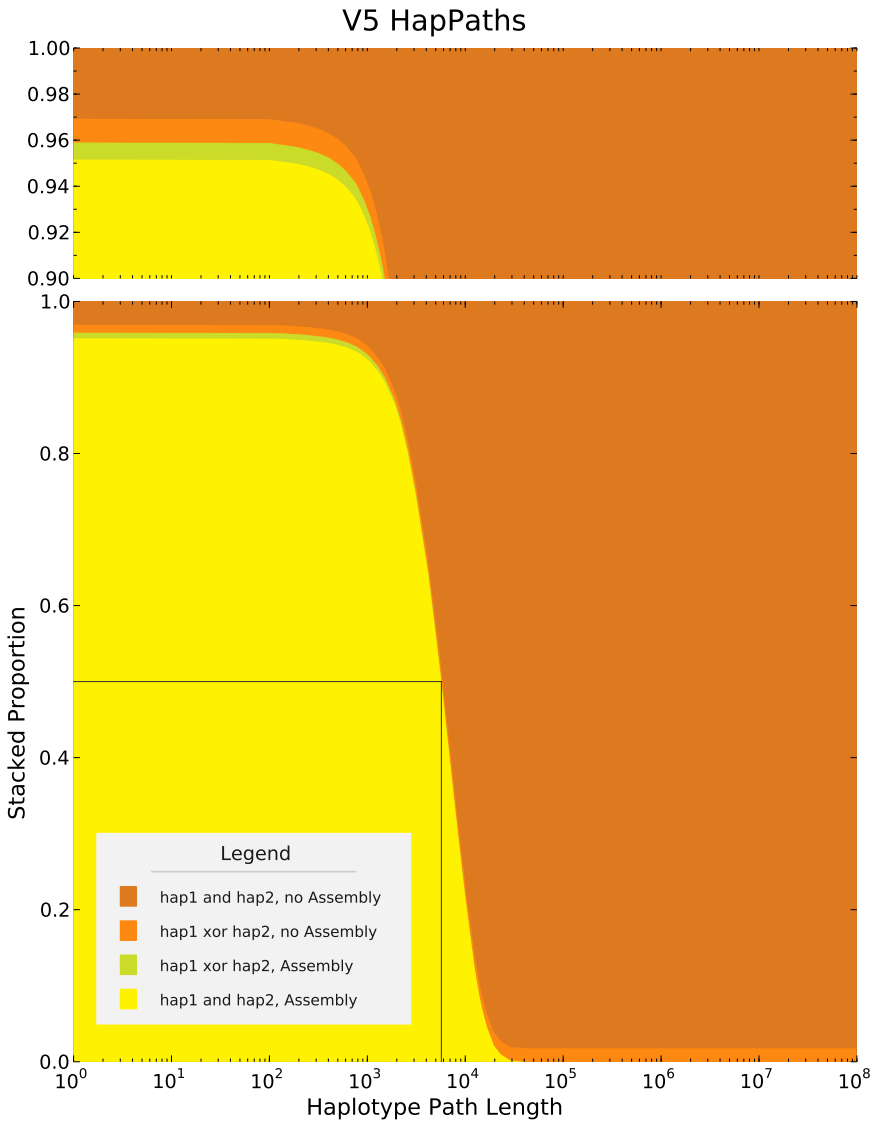


Figure 3.189: V5 hapPaths caption goes here.

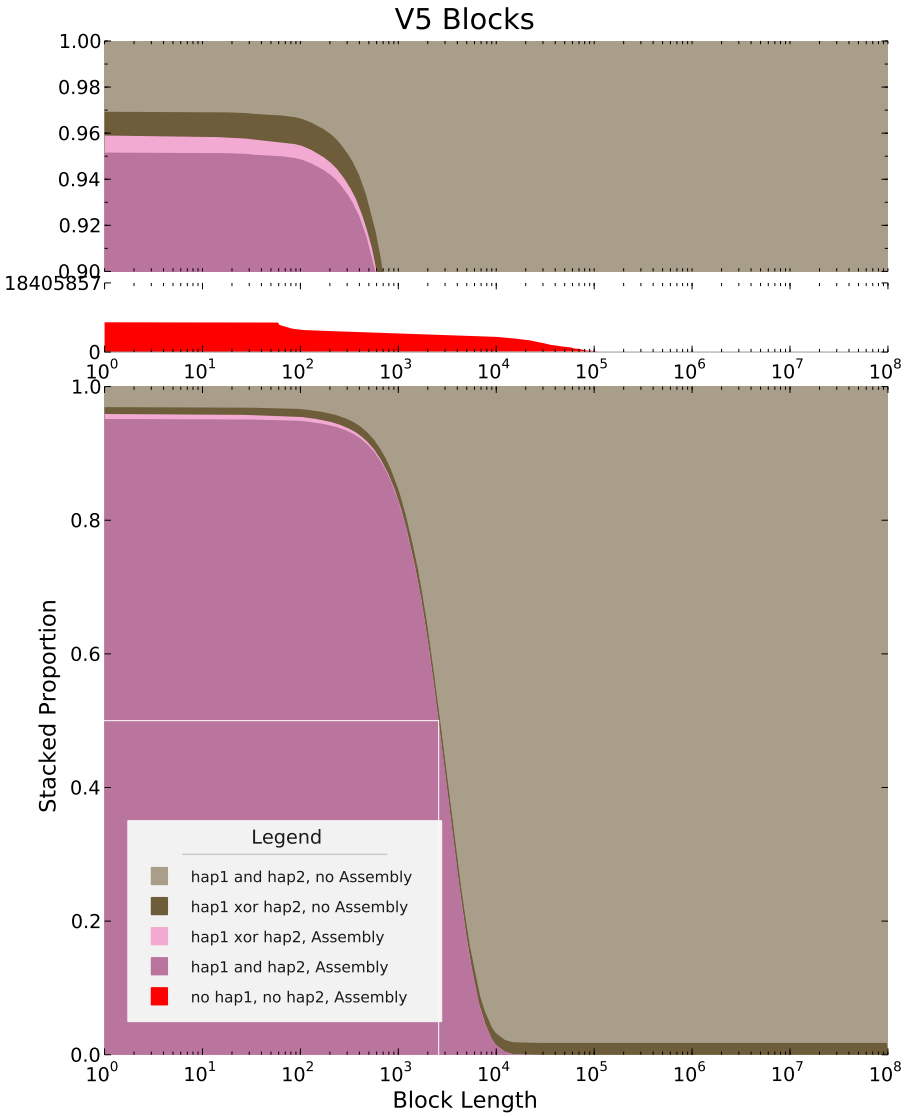


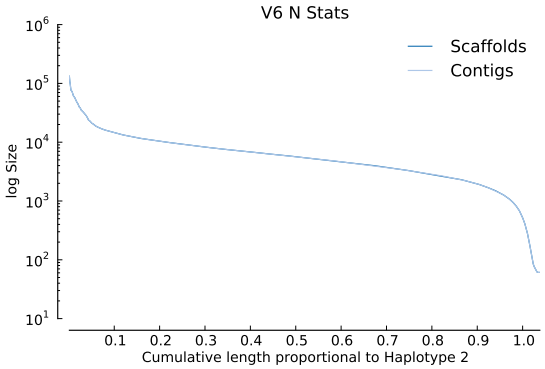
Figure 3.190: V5 blocks caption goes here.

V6

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
V5	0.96373	0.96390	0.96357	0.99559
V6	0.96157	0.96183	0.96132	0.99560
V4	0.96153	0.96180	0.96128	0.99789

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	65,820	61	69.00	220	1,774.43	2,426.00	132,122	3,324.77	116,793,074
Contigs	65,834	61	69.00	220	1,774.05	2,426.75	132,122	3,323.54	116,792,643

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,920,310 – 110,354,787	105,769,637 – 106,667,798	211,528,070.0 – 213,302,438.0	4,598 – 7,745
Heterozygous	419,615 – 439,603	404,781 – 415,701	809,462.0 – 830,846.0	16 – 23
Indel	1,451,820 – 1,841,963	560,946 – 735,902	1,120,122.0 – 1,467,246.0	864 – 1,273

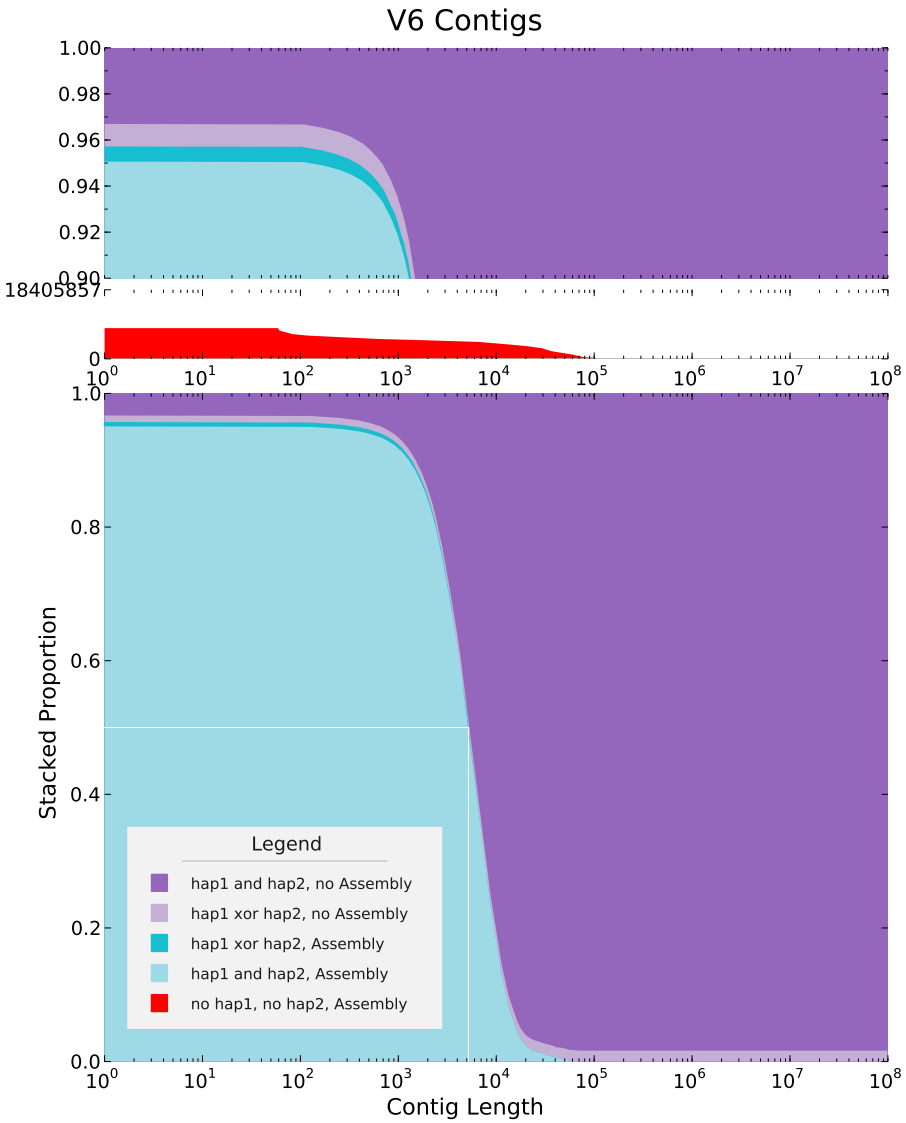


Figure 3.191: V6 contigs caption goes here.

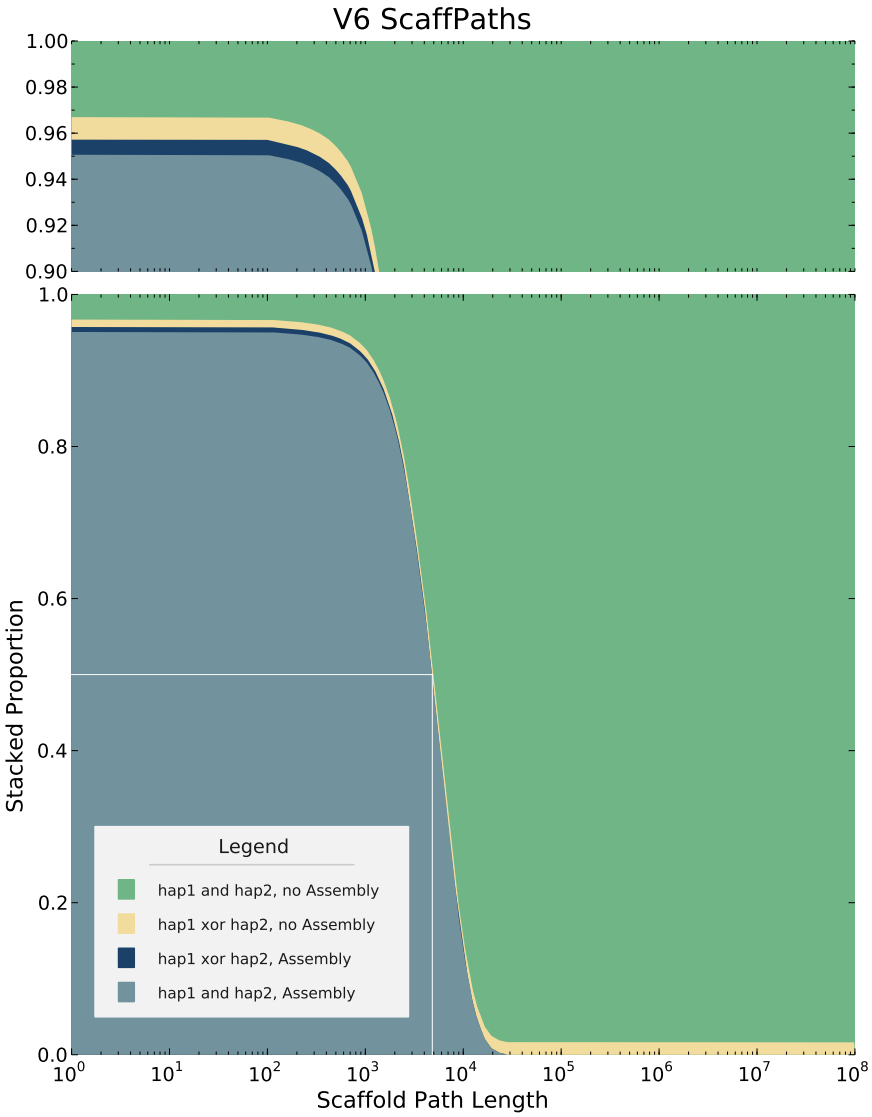


Figure 3.192: V6 scaffolds caption goes here.

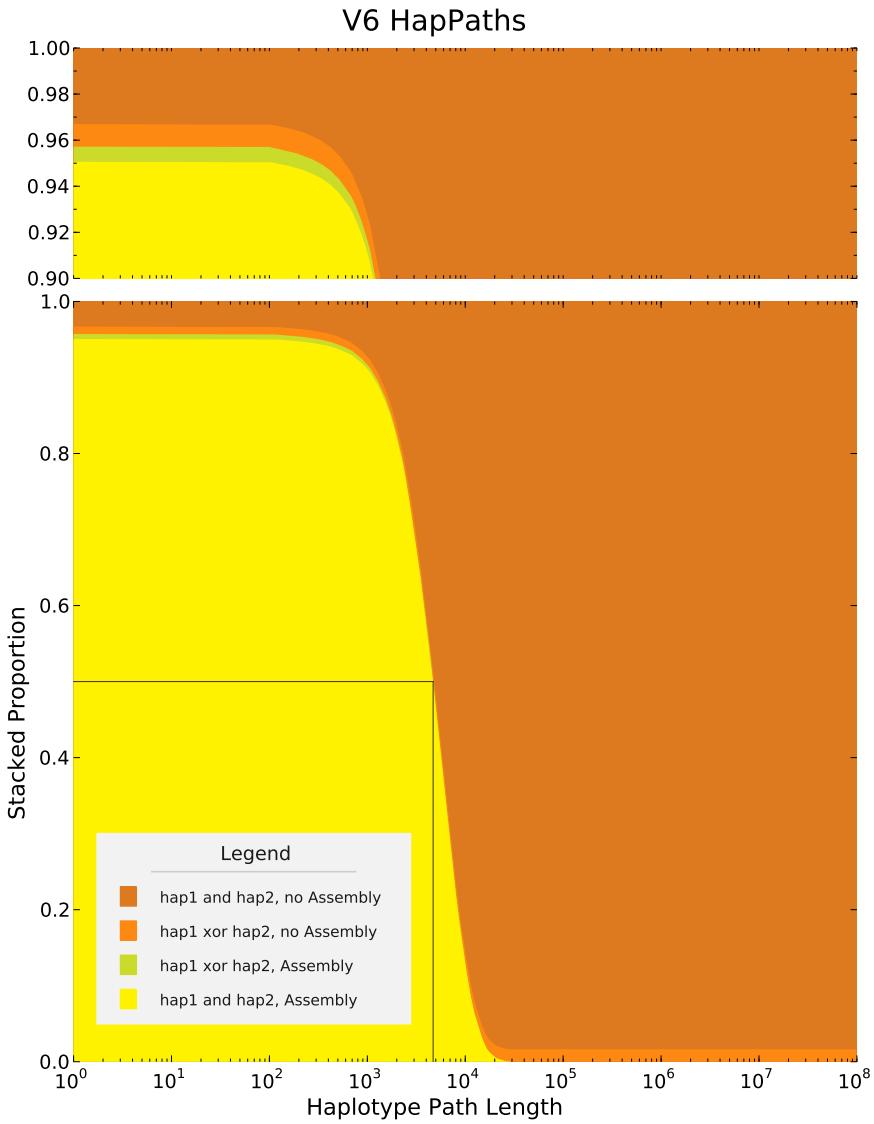


Figure 3.193: V6 hapPaths caption goes here.

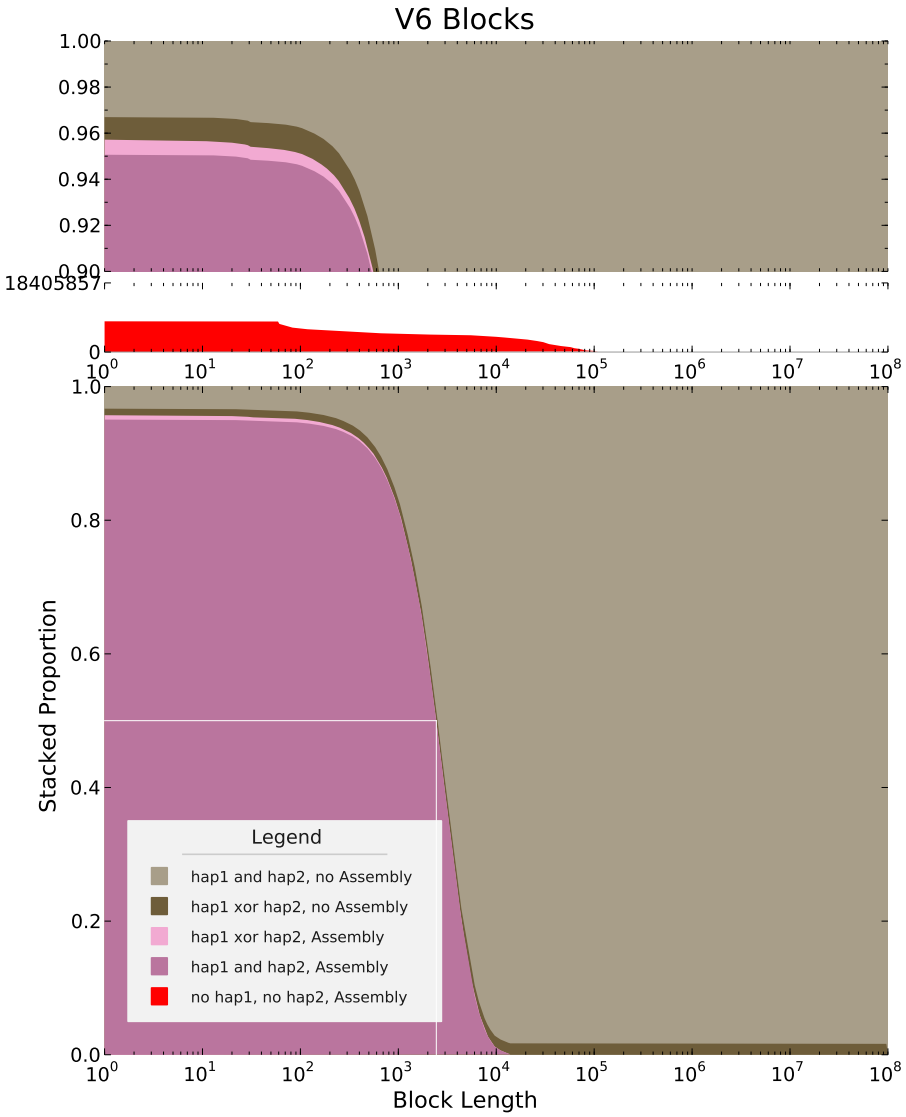


Figure 3.194: V6 blocks caption goes here.

3.2.19 W, Auto

Affiliation: Auto
Contact: Auto
Software: **CLC**
Number of entries: 11

ID	Total	Hap 1	Hap 2	Bac
W8	0.97204	0.97232	0.97175	0.99831
W11	0.97203	0.97220	0.97187	0.99798
W5	0.97126	0.97152	0.97102	0.99800
W9	0.97080	0.97102	0.97057	0.99805
W1	0.97034	0.97048	0.97023	0.99825
W7	0.96984	0.97006	0.96961	0.99806
W6	0.96892	0.96918	0.96865	0.99764
W10	0.96812	0.96852	0.96772	0.99812
W3	0.96751	0.96771	0.96728	0.99810

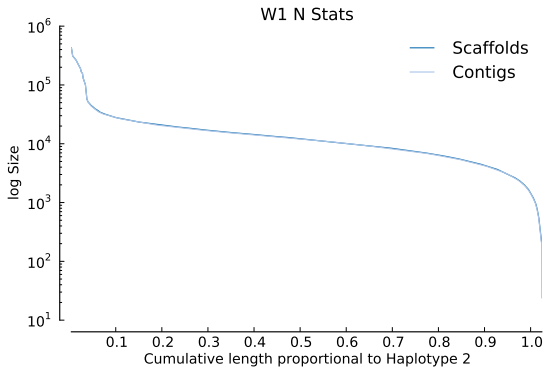
Assemblies:

W1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
W9	0.97080	0.97102	0.97057	0.99805
W1	0.97034	0.97048	0.97023	0.99825
M4	0.97031	0.97047	0.97014	0.99718

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	17,187	200	1,398.00	4,444	6,706.47	9,549.50	428,217	9,751.55	115,264,020
Contigs	17,759	24	1,252.00	4,237	6,489.44	9,267.50	428,217	9,604.57	115,245,941

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,163,268 – 110,224,825	106,909,158 – 107,648,097	213,786,064.0 – 215,244,720.0	1,551 – 4,294
Heterozygous	423,544 – 439,925	412,374 – 420,218	824,648.0 – 840,216.0	10 – 46
Indel	2,075,651 – 2,453,341	905,380 – 1,058,778	1,808,046.0 – 2,113,630.0	1,226 – 1,472

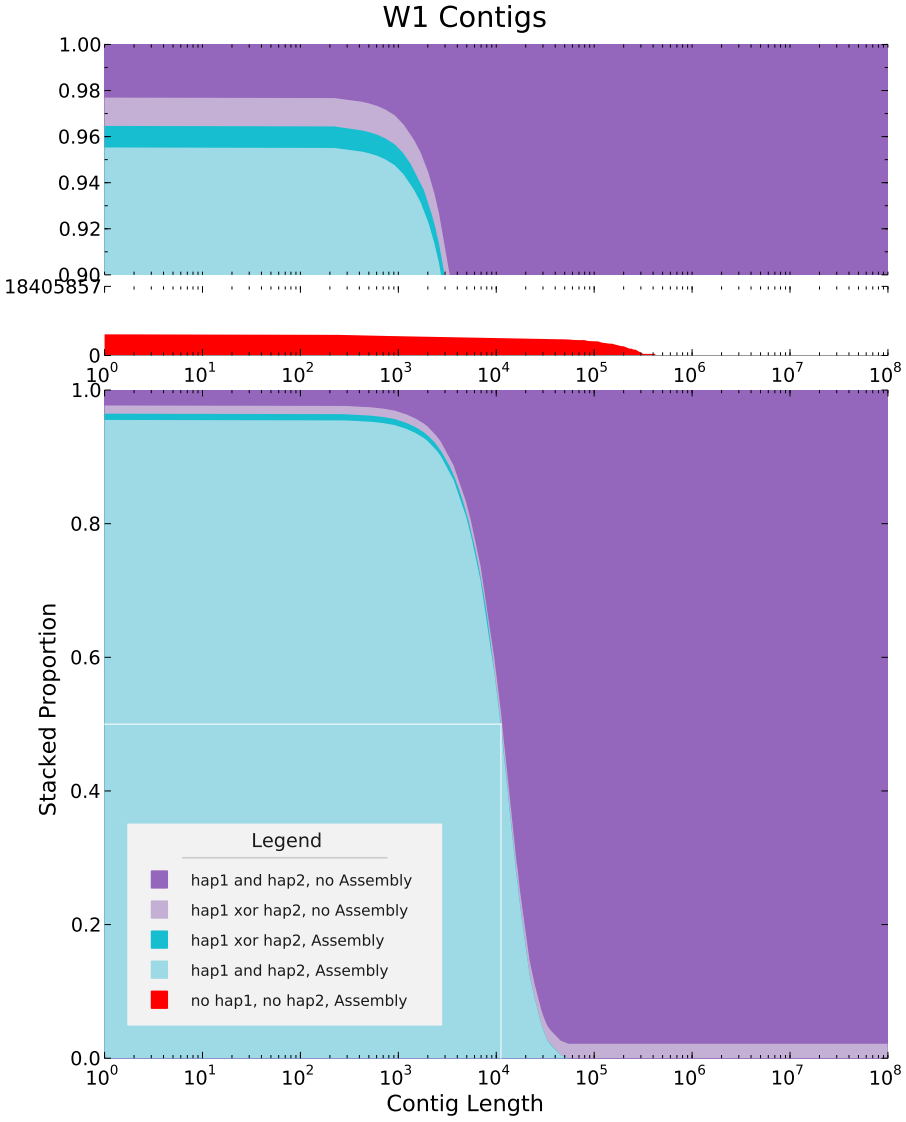


Figure 3.195: W1 contigs caption goes here.

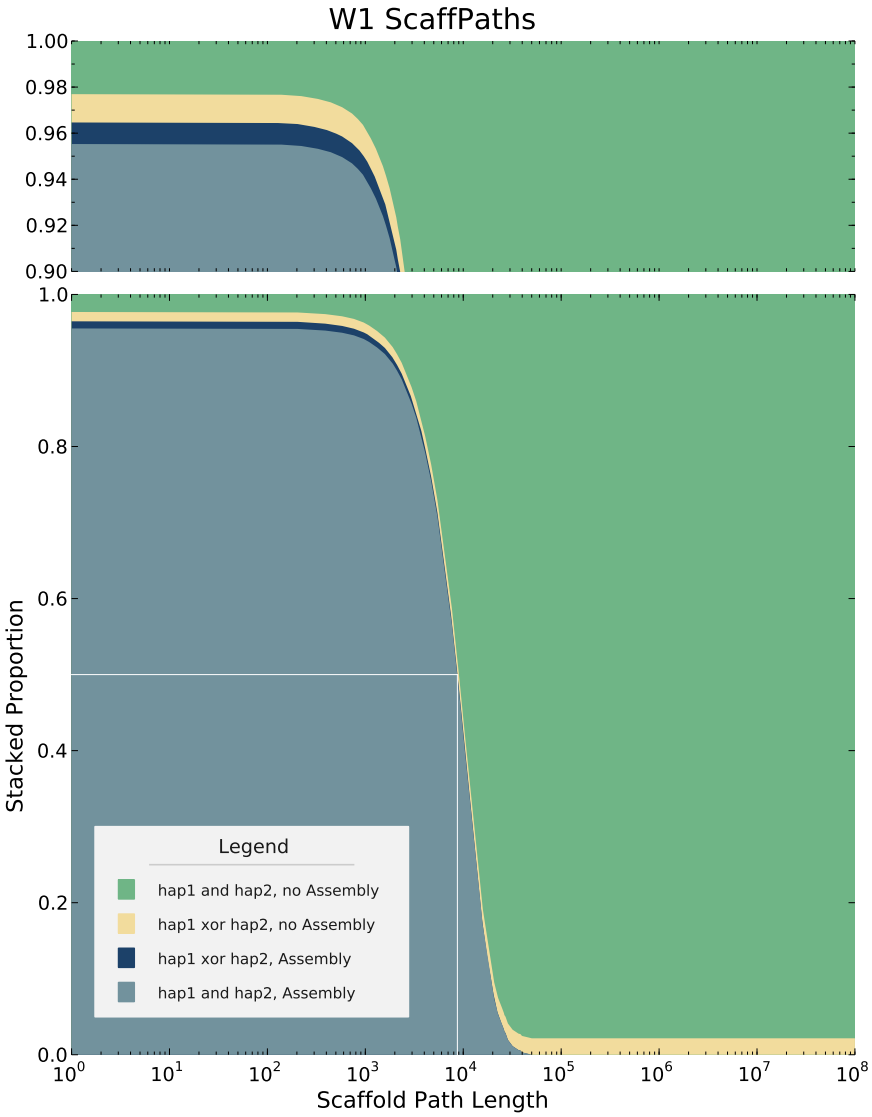


Figure 3.196: W1 scaffolds caption goes here.

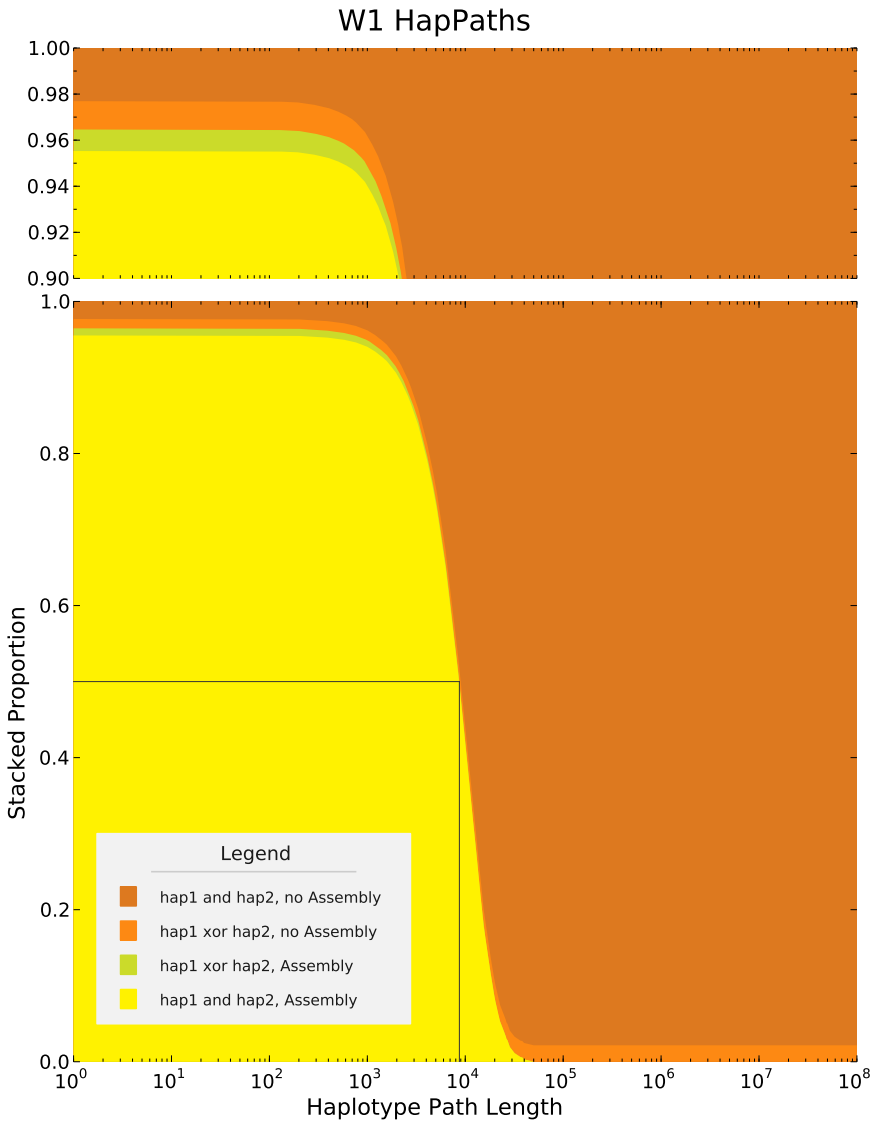


Figure 3.197: W1 hapPaths caption goes here.

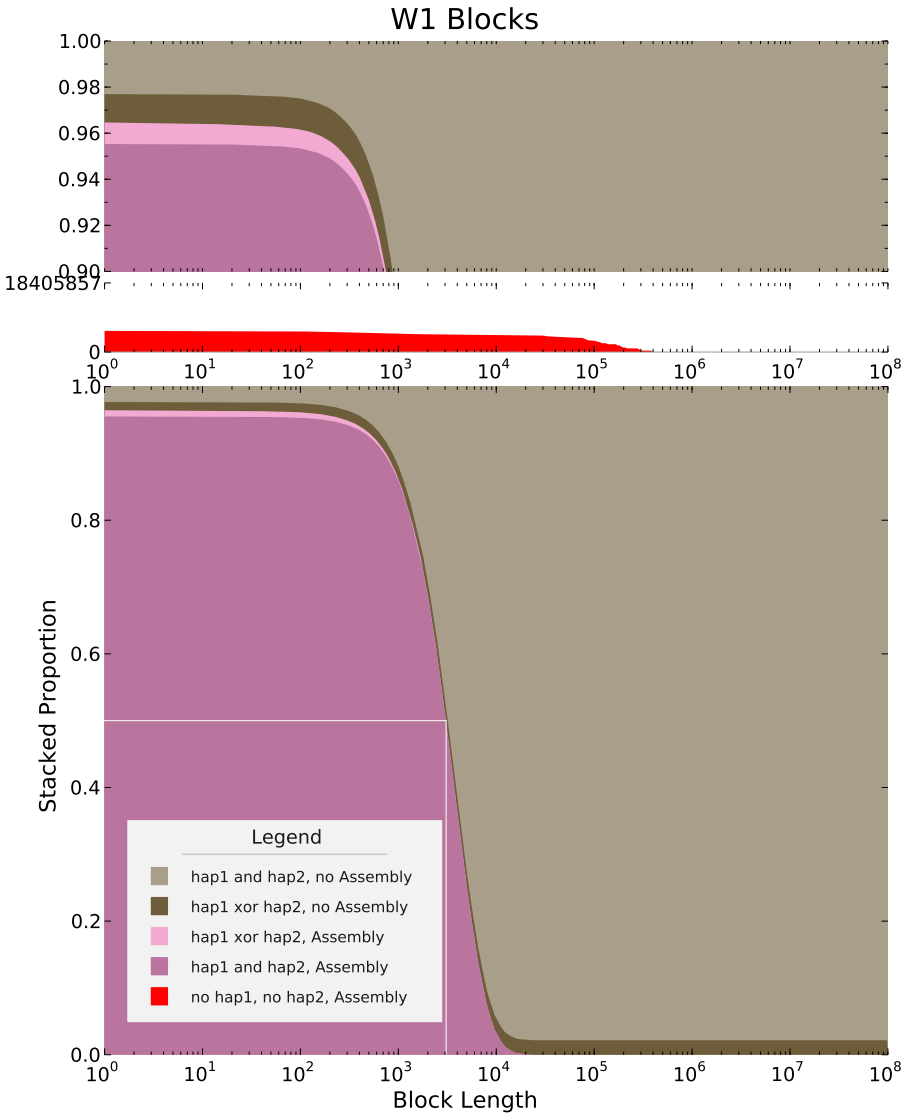


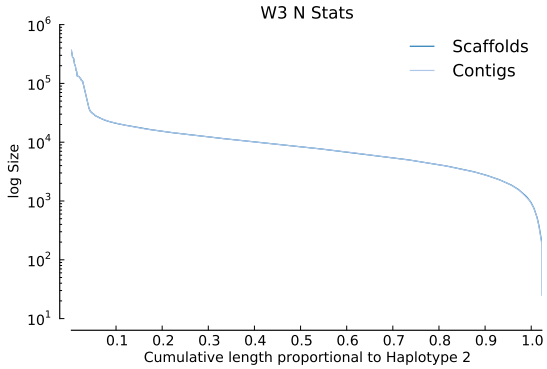
Figure 3.198: W1 blocks caption goes here.

W3

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
X4	0.96758	0.96779	0.96735	0.99798
W3	0.96751	0.96771	0.96728	0.99810
V5	0.96373	0.96390	0.96357	0.99559

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	24,858	200	1,079.00	2,990	4,630.87	6,249.00	364,958	6,930.83	115,114,064
Contigs	25,328	24	1,001.00	2,898	4,544.33	6,150.25	364,958	6,887.58	115,098,816

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,146,021 – 110,312,968	106,667,991 – 107,413,680	213,319,492.0 – 214,791,202.0	1,361 – 4,127
Heterozygous	419,853 – 439,919	407,565 – 415,051	815,054.0 – 829,906.0	15 – 56
Indel	1,809,799 – 2,184,873	762,257 – 887,975	1,522,216.0 – 1,772,740.0	1,089 – 1,380

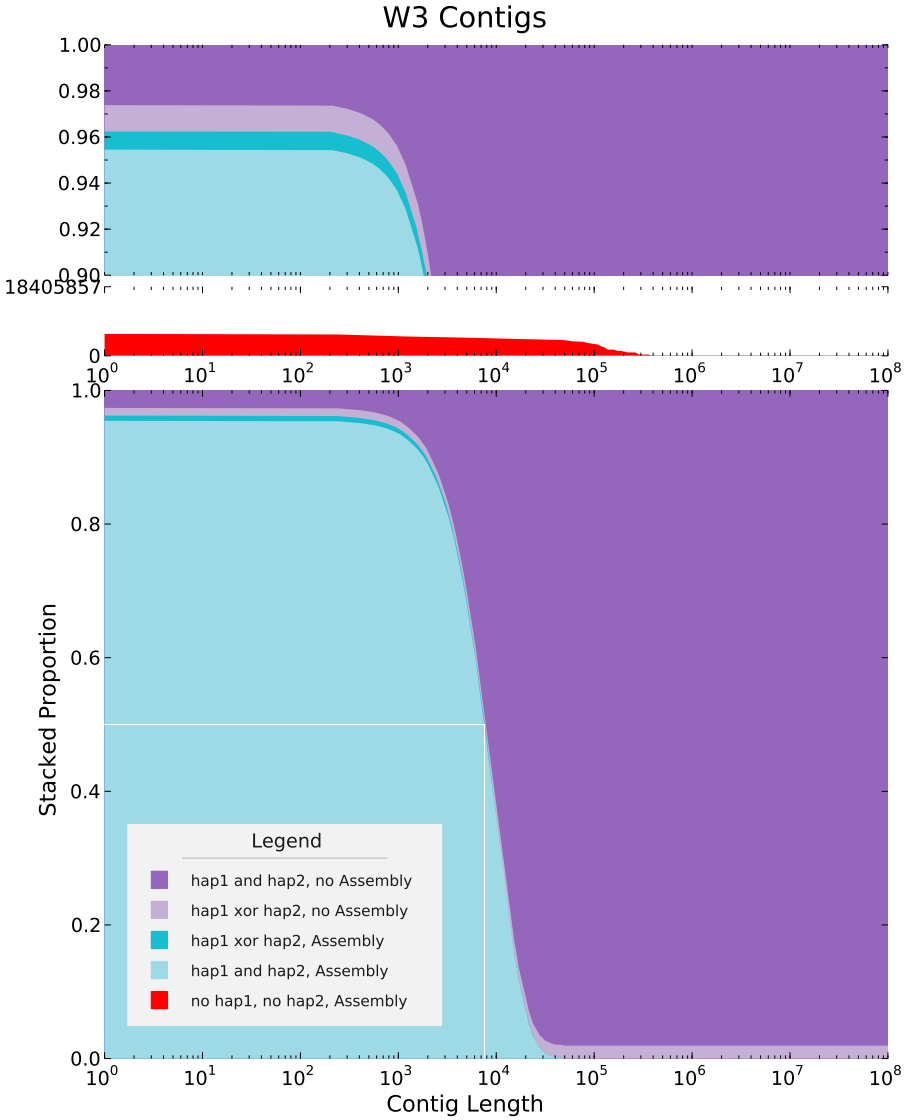


Figure 3.199: W3 contigs caption goes here.

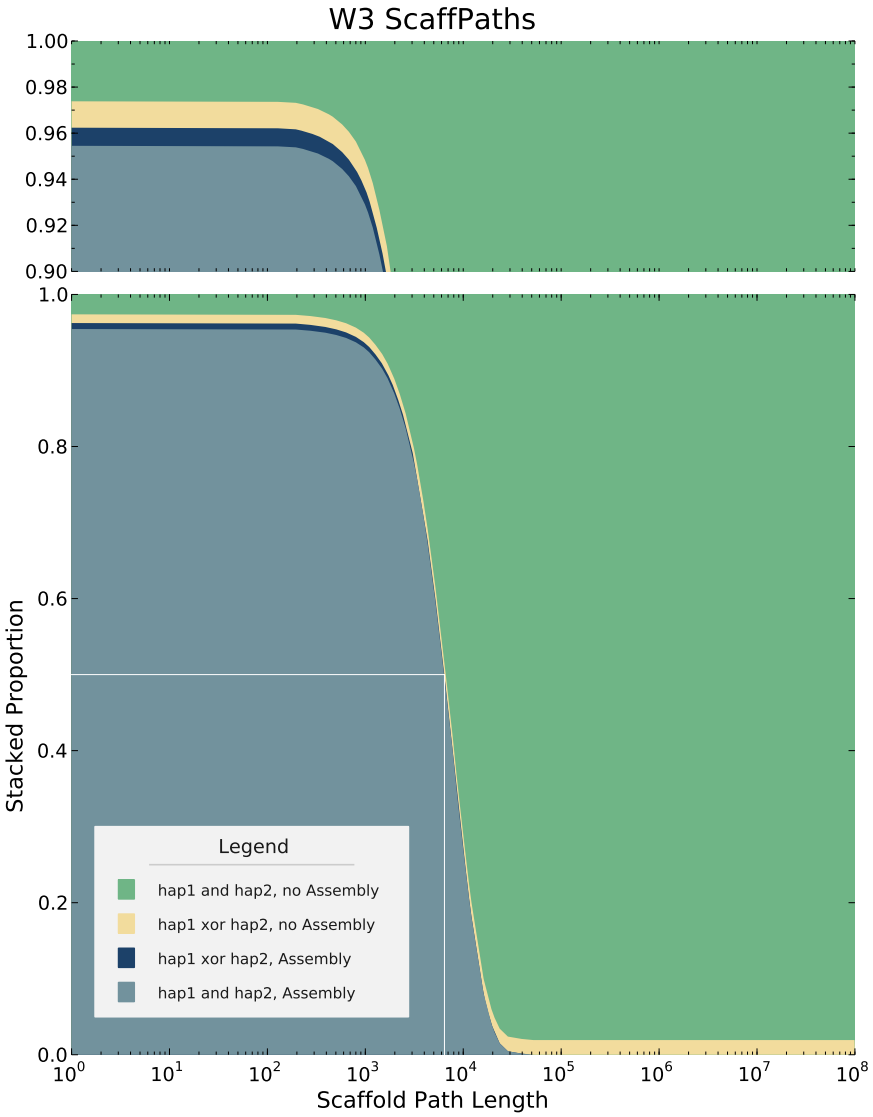


Figure 3.200: W3 scaffolds caption goes here.

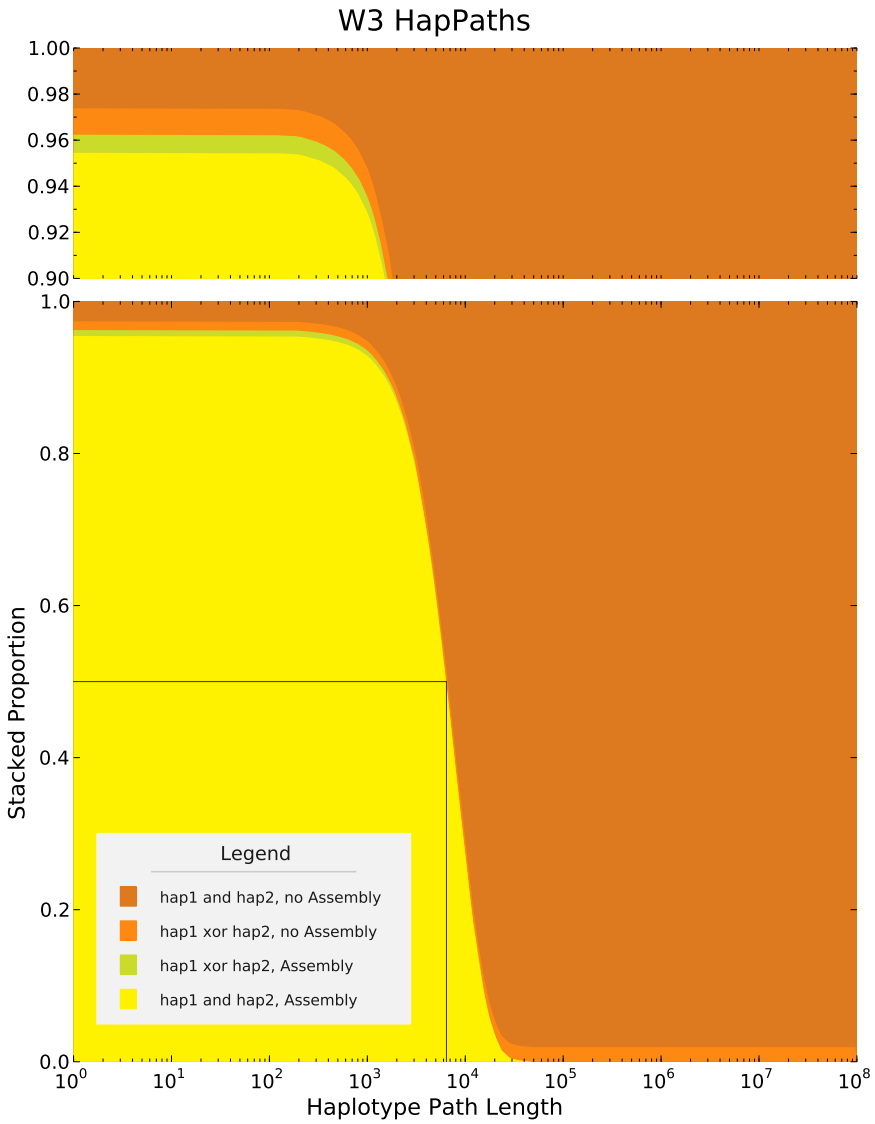


Figure 3.201: W3 hapPaths caption goes here.

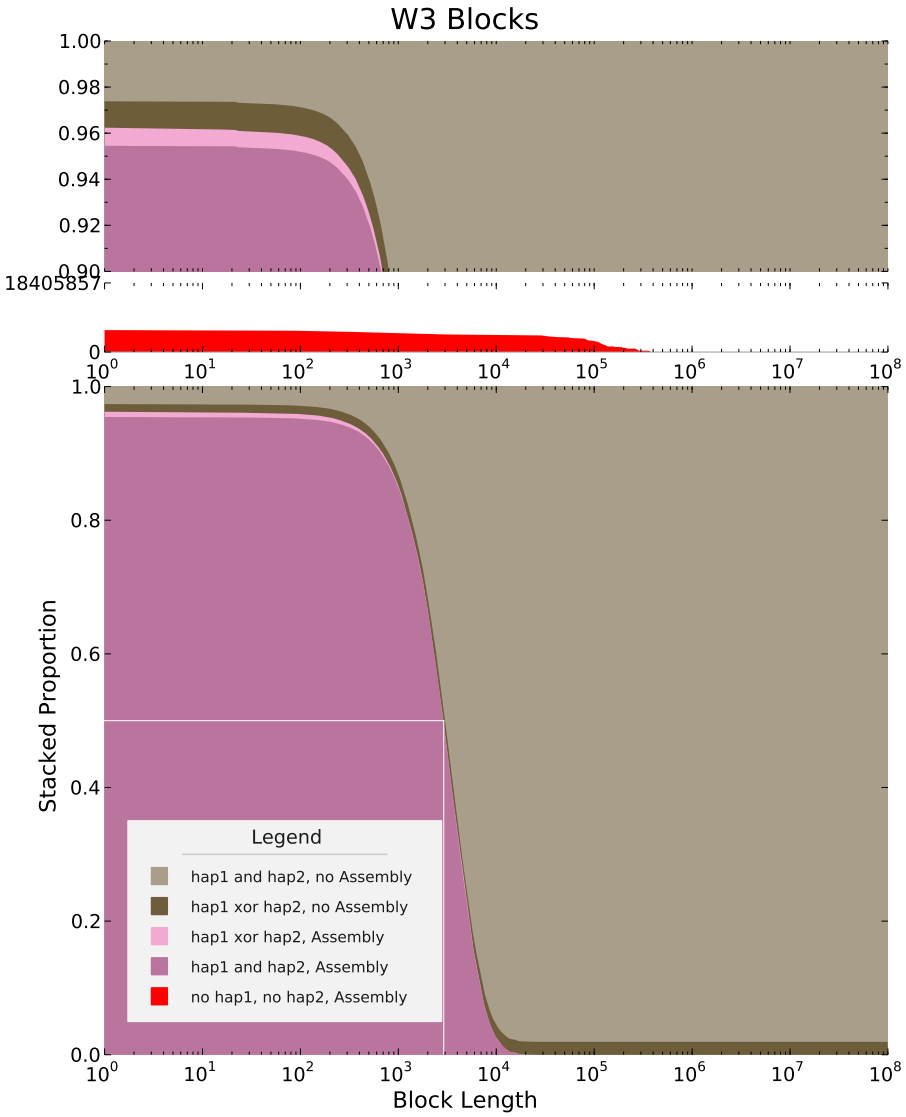


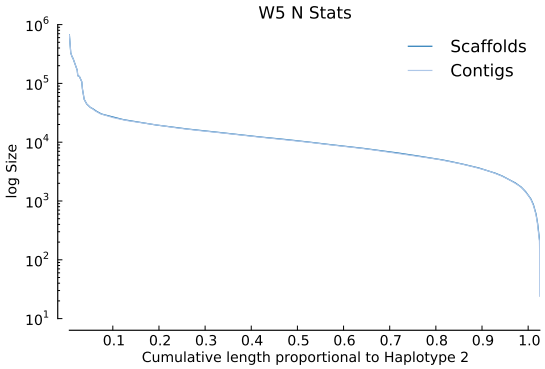
Figure 3.202: W3 blocks caption goes here.

W5

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
W11	0.97203	0.97220	0.97187	0.99798
W5	0.97126	0.97152	0.97102	0.99800
W9	0.97080	0.97102	0.97057	0.99805

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	19,928	200	1,353.75	3,704	5,789.54	7,799.00	665,472	9,352.26	115,373,922
Contigs	20,561	24	1,217.00	3,533	5,610.37	7,566.00	665,472	9,237.28	115,354,785

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,123,615 – 110,209,848	107,044,765 – 107,765,270	214,062,756.0 – 215,485,536.0	1,562 – 4,113
Heterozygous	420,117 – 439,815	409,688 – 416,978	819,292.0 – 833,762.0	11 – 37
Indel	2,070,348 – 2,447,854	895,766 – 1,028,800	1,788,828.0 – 2,054,078.0	1,268 – 1,467

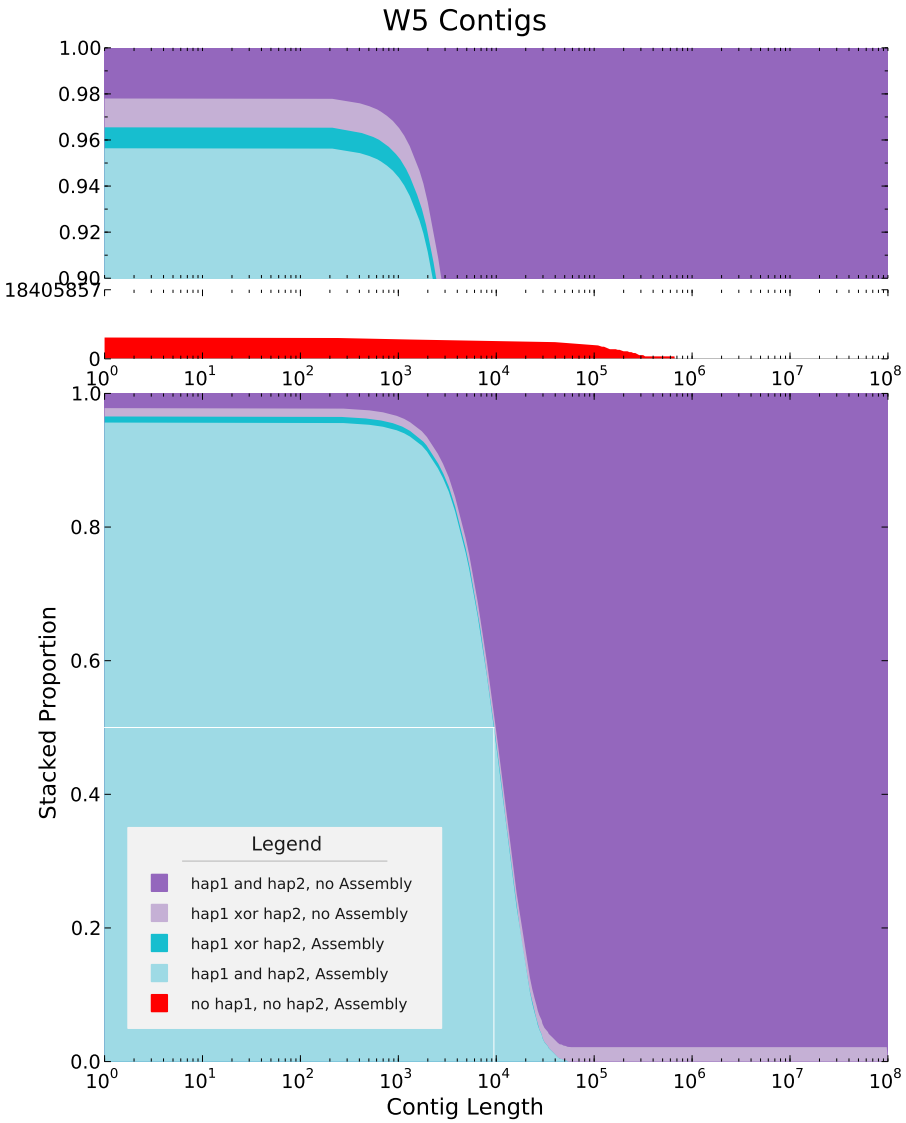


Figure 3.203: W5 contigs caption goes here.

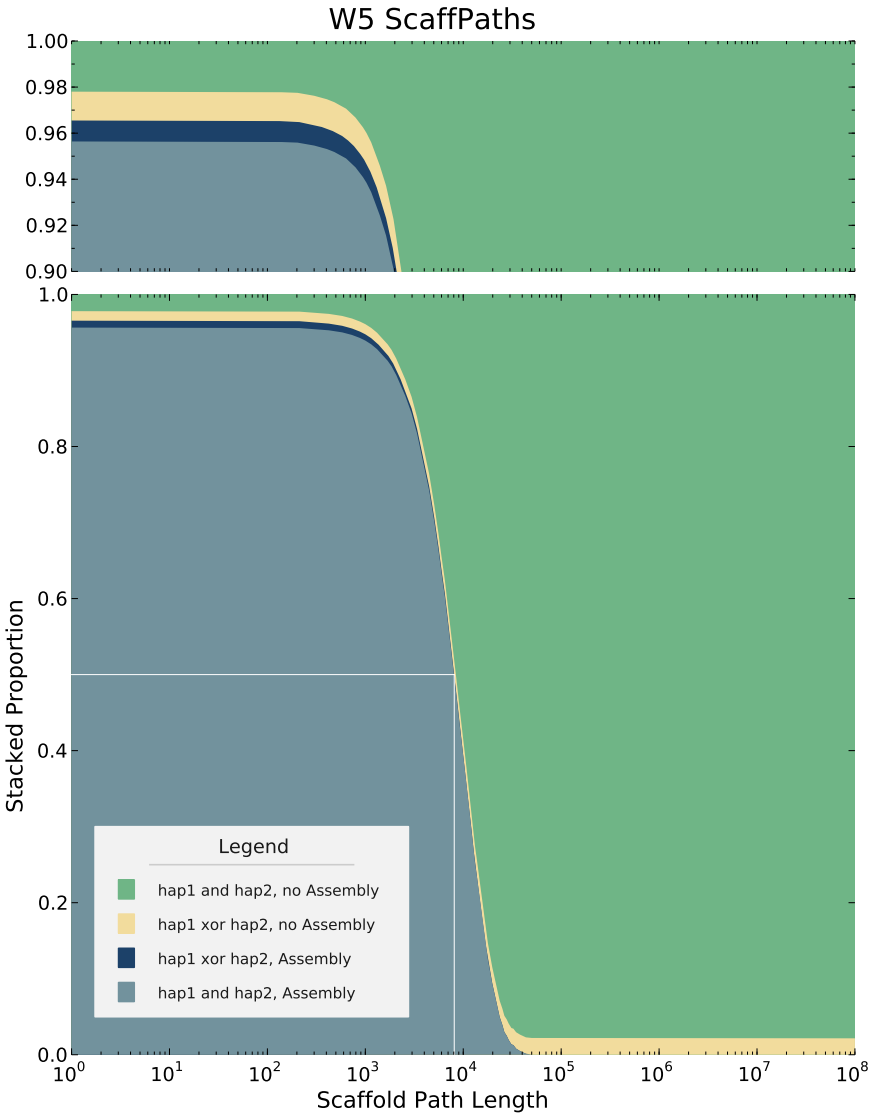


Figure 3.204: W5 scaffolds caption goes here.

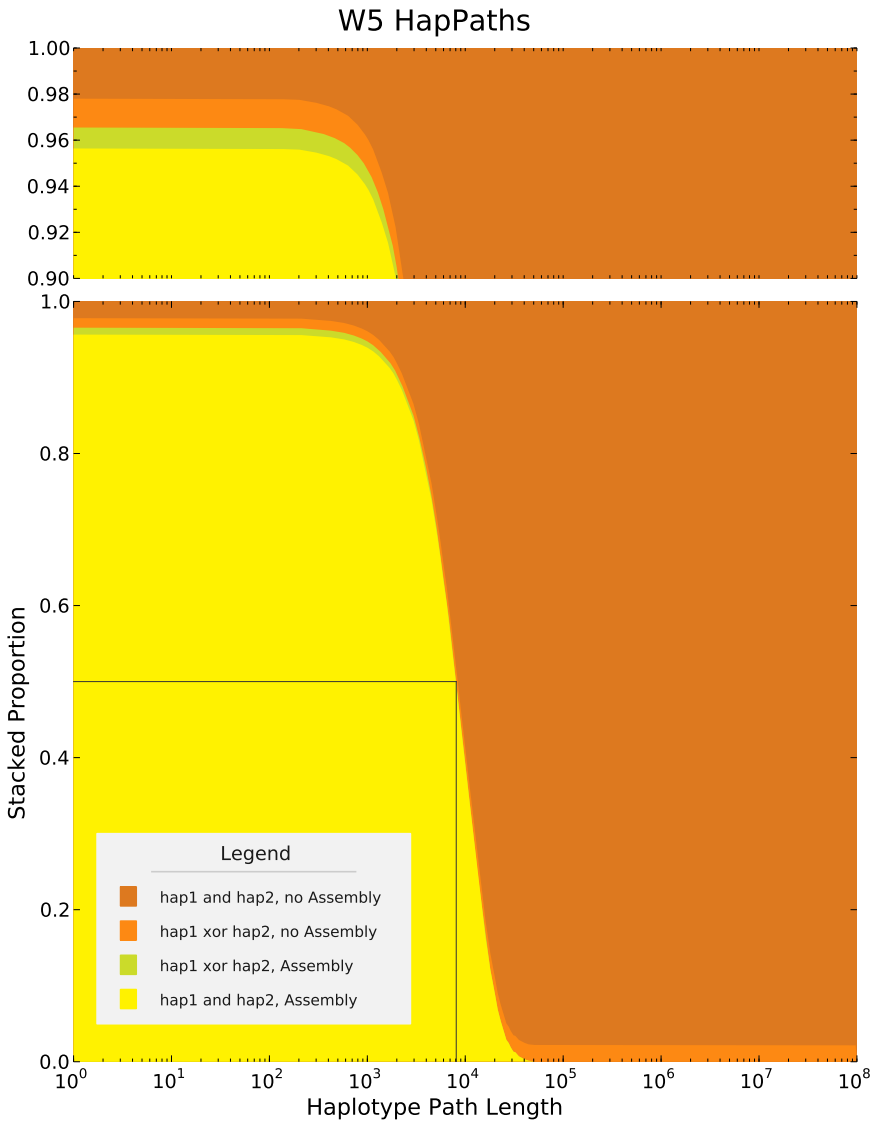


Figure 3.205: W5 hapPaths caption goes here.

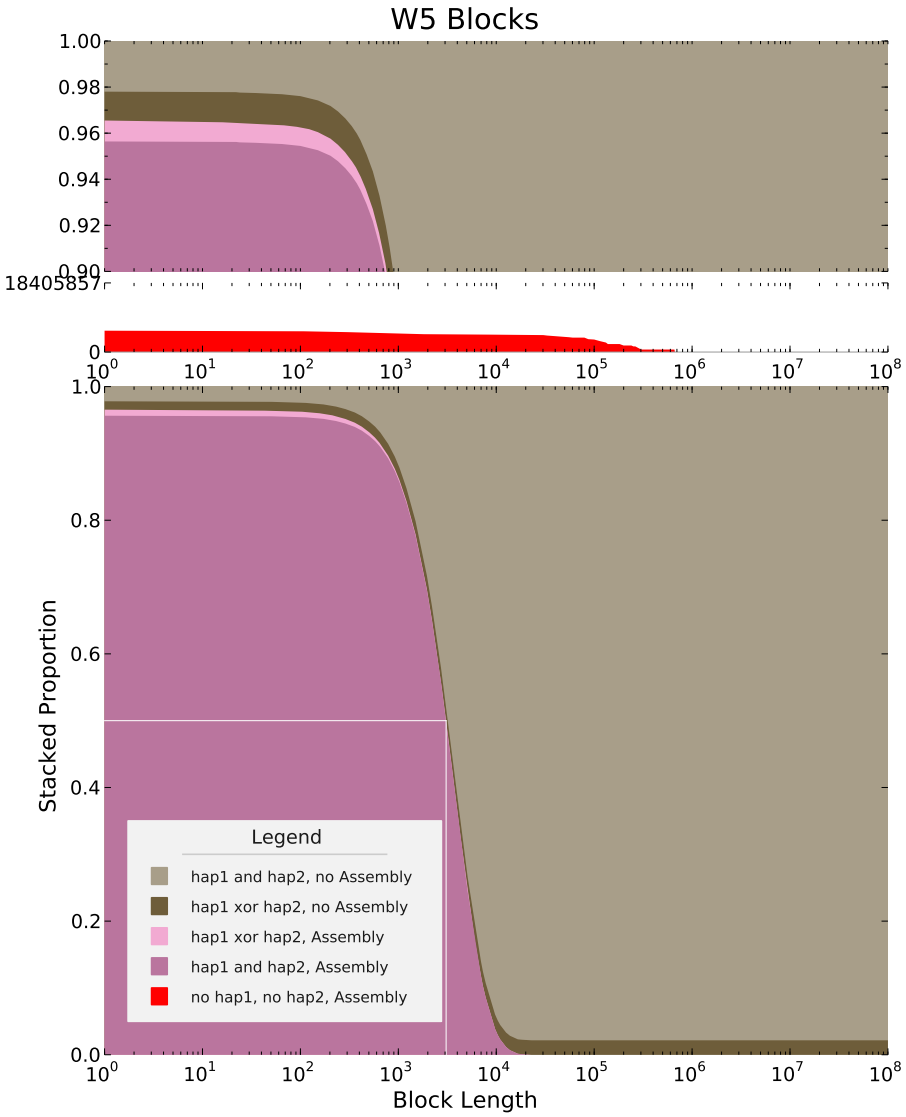


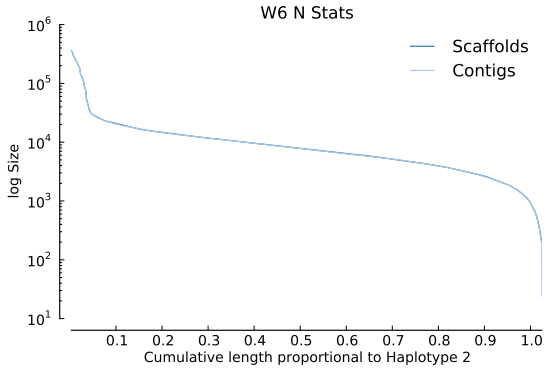
Figure 3.206: W5 blocks caption goes here.

W6

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
G1	0.96922	0.96951	0.96894	0.99498
W6	0.96892	0.96918	0.96865	0.99764
W10	0.96812	0.96852	0.96772	0.99812

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	25,991	200	1,071.50	2,845	4,436.04	5,946.50	364,958	7,112.35	115,297,130
Contigs	26,531	24	989.00	2,768	4,345.14	5,858.00	364,958	7,056.25	115,280,793

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,080,824 – 110,244,960	106,810,762 – 107,547,714	213,601,756.0 – 215,057,646.0	1,654 – 4,000
Heterozygous	416,281 – 439,697	404,995 – 412,381	809,934.0 – 824,596.0	7 – 29
Indel	1,958,719 – 2,334,254	822,114 – 946,549	1,641,882.0 – 1,890,034.0	1,137 – 1,334

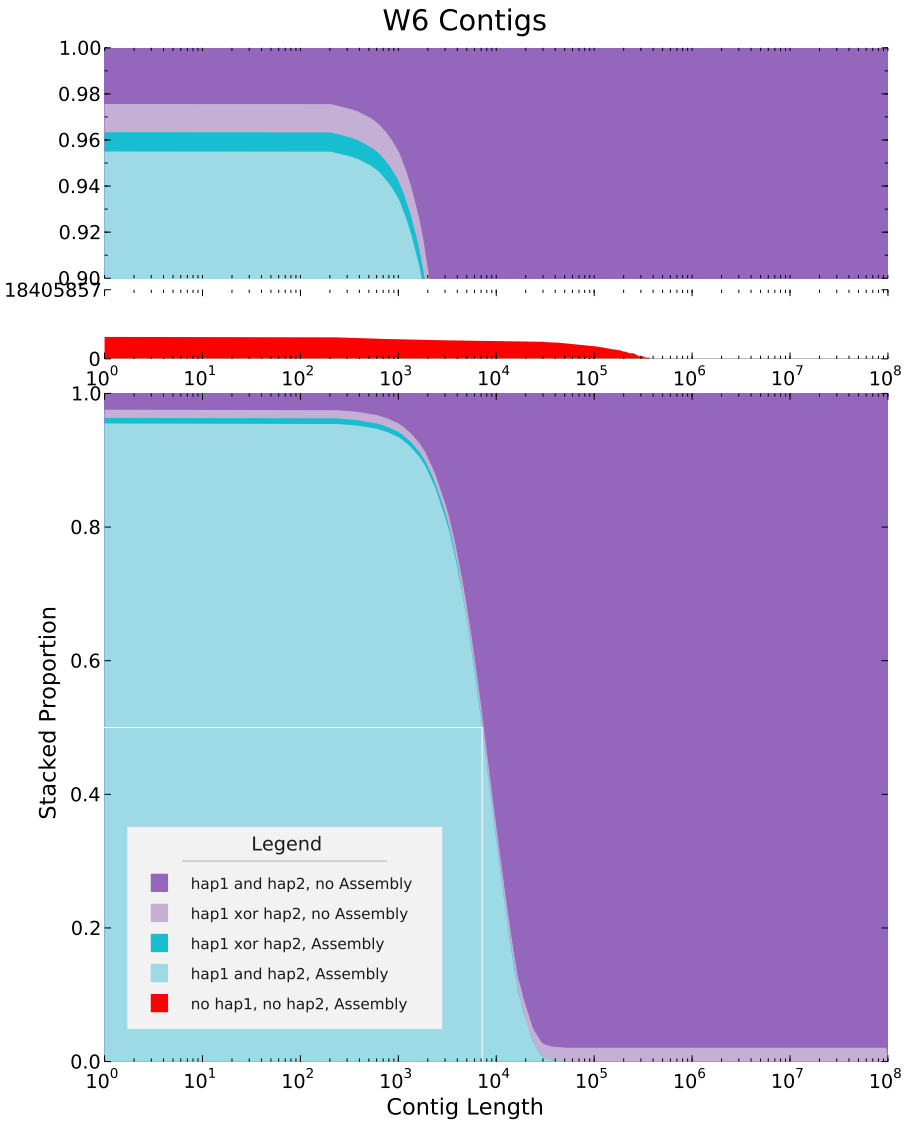


Figure 3.207: W6 contigs caption goes here.

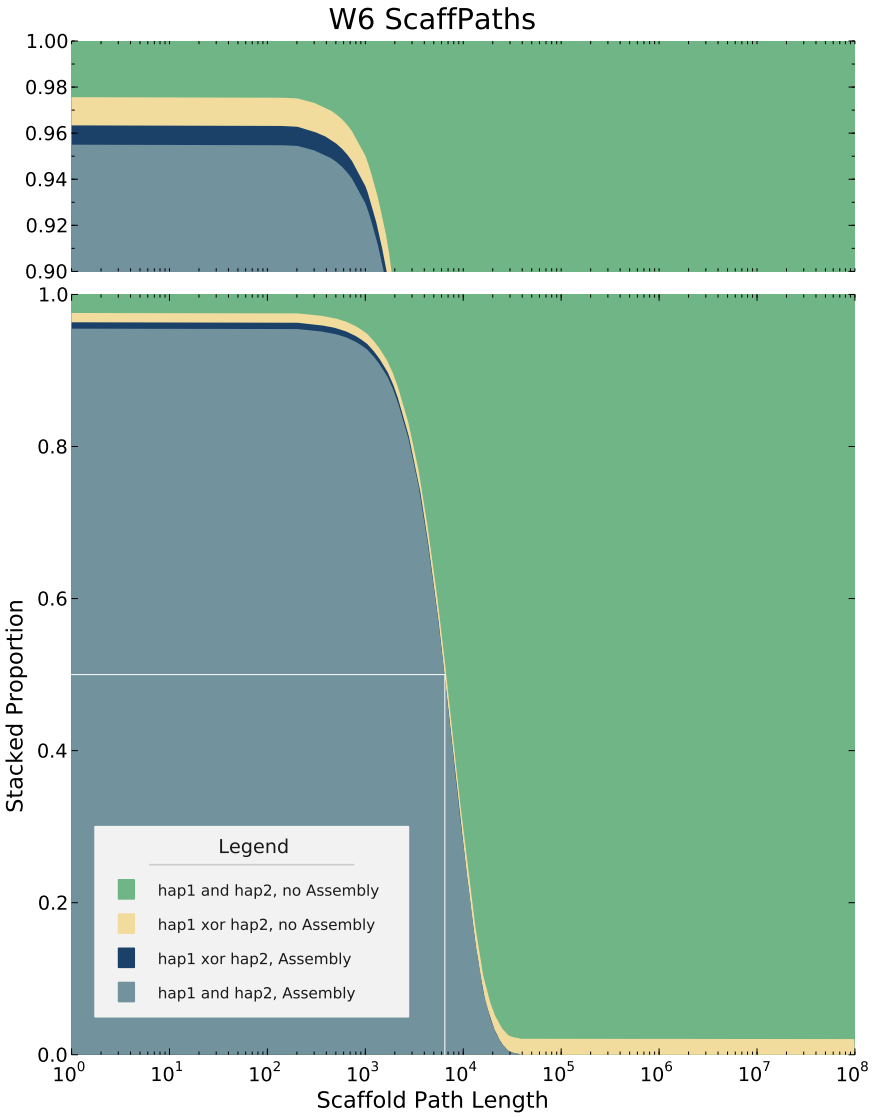


Figure 3.208: W6 scaffolds caption goes here.

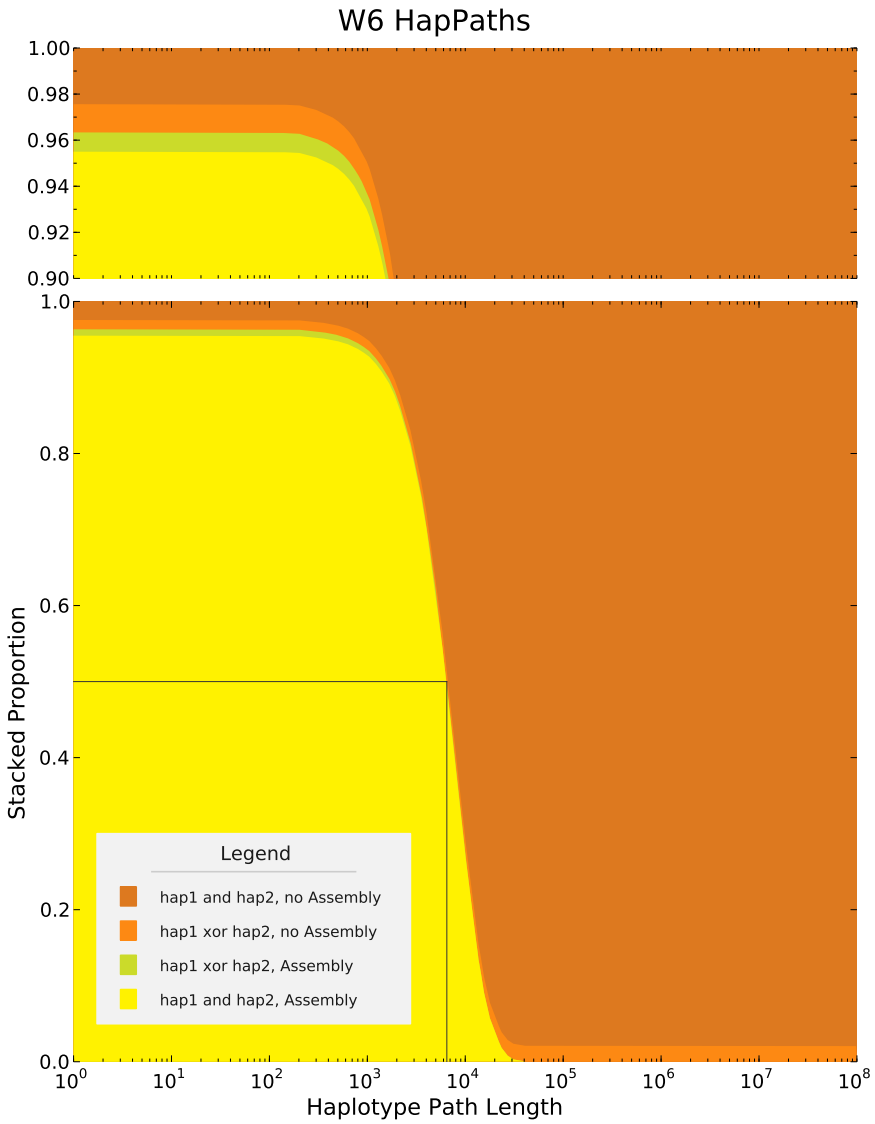


Figure 3.209: W6 hapPaths caption goes here.

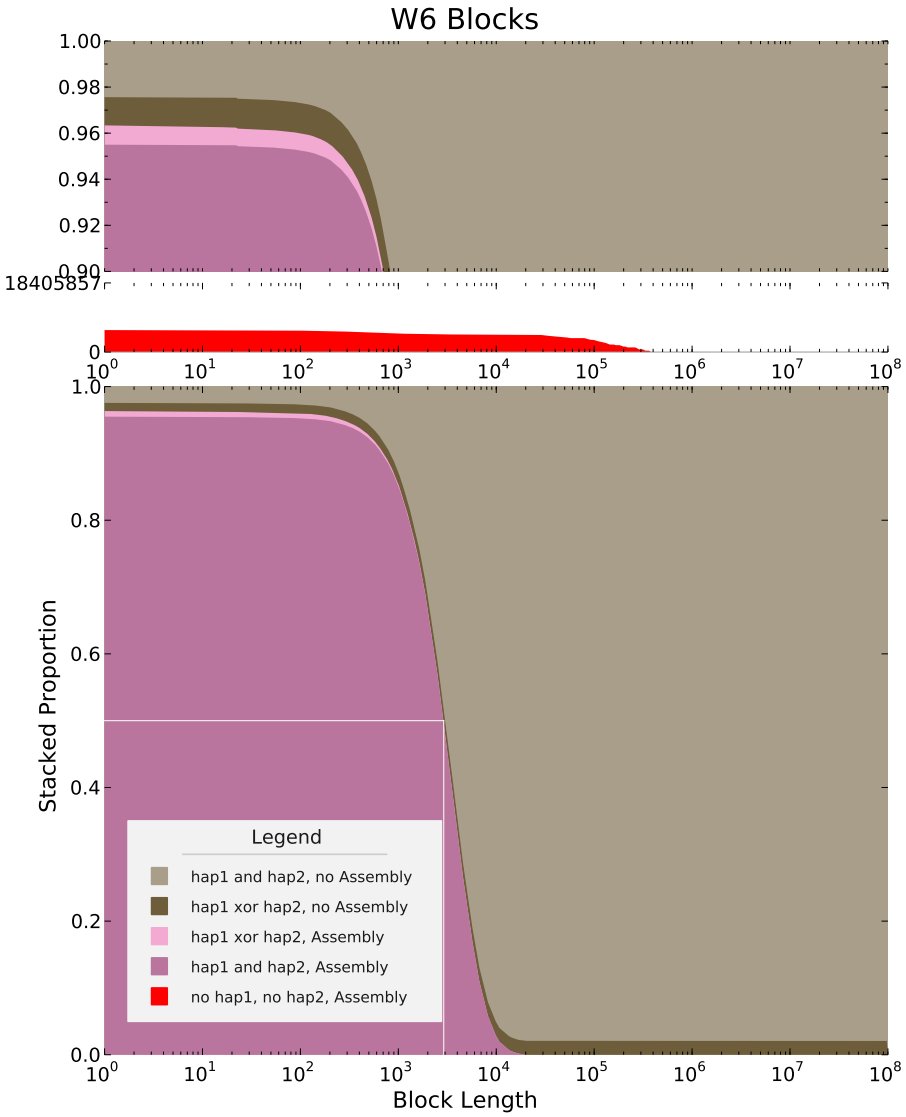


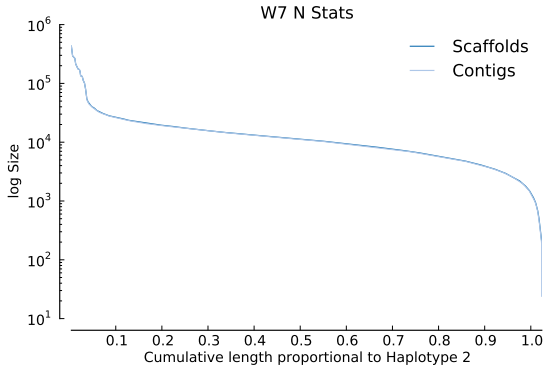
Figure 3.210: W6 blocks caption goes here.

W7

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
M4	0.97031	0.97047	0.97014	0.99718
W7	0.96984	0.97006	0.96961	0.99806
G1	0.96922	0.96951	0.96894	0.99498

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	18,342	200	1,439.25	4,192	6,281.64	8,814.25	428,202	8,943.59	115,217,841
Contigs	18,929	24	1,271.00	3,992	6,085.85	8,580.00	428,202	8,826.32	115,199,002

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,177,431 – 110,257,267	106,877,032 – 107,609,778	213,724,318.0 – 215,170,768.0	1,390 – 4,193
Heterozygous	423,368 – 440,084	411,875 – 419,496	823,648.0 – 838,770.0	8 – 44
Indel	2,005,360 – 2,382,446	877,397 – 1,029,185	1,752,130.0 – 2,054,466.0	1,170 – 1,382

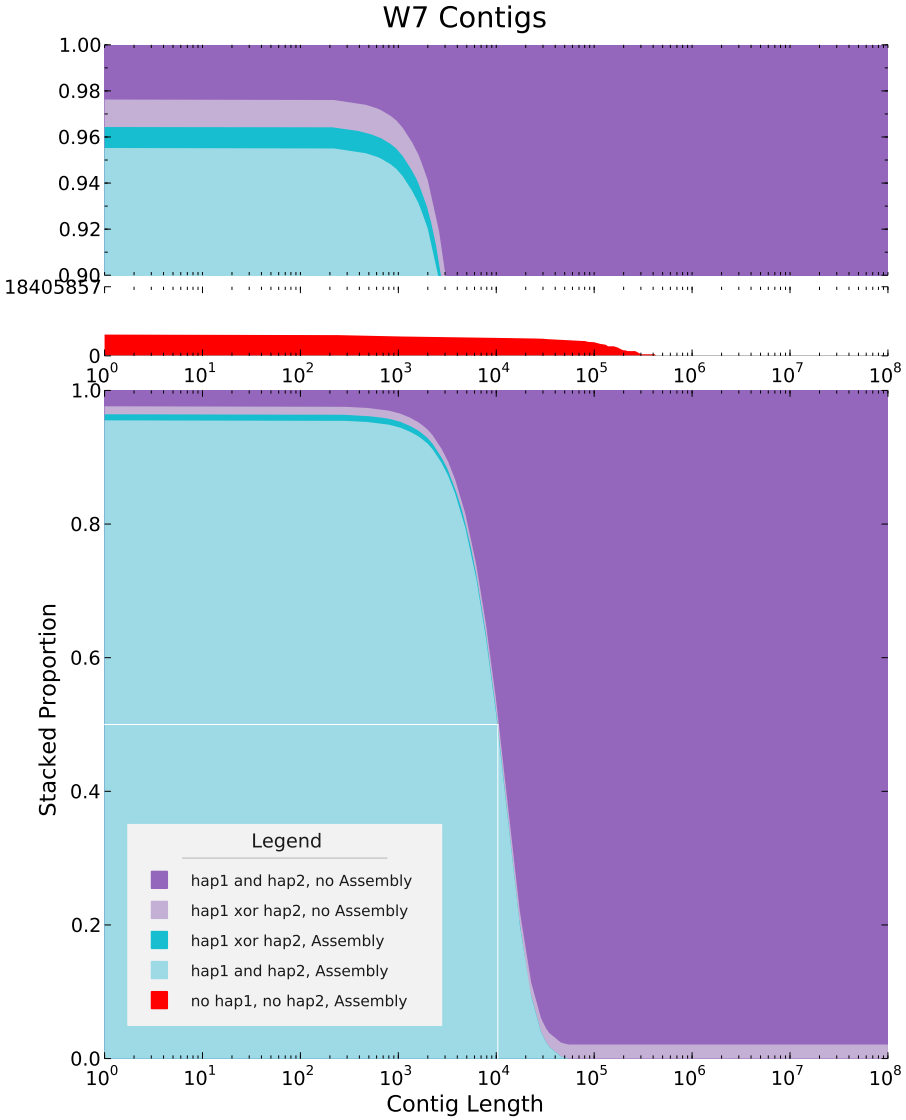


Figure 3.211: W7 contigs caption goes here.

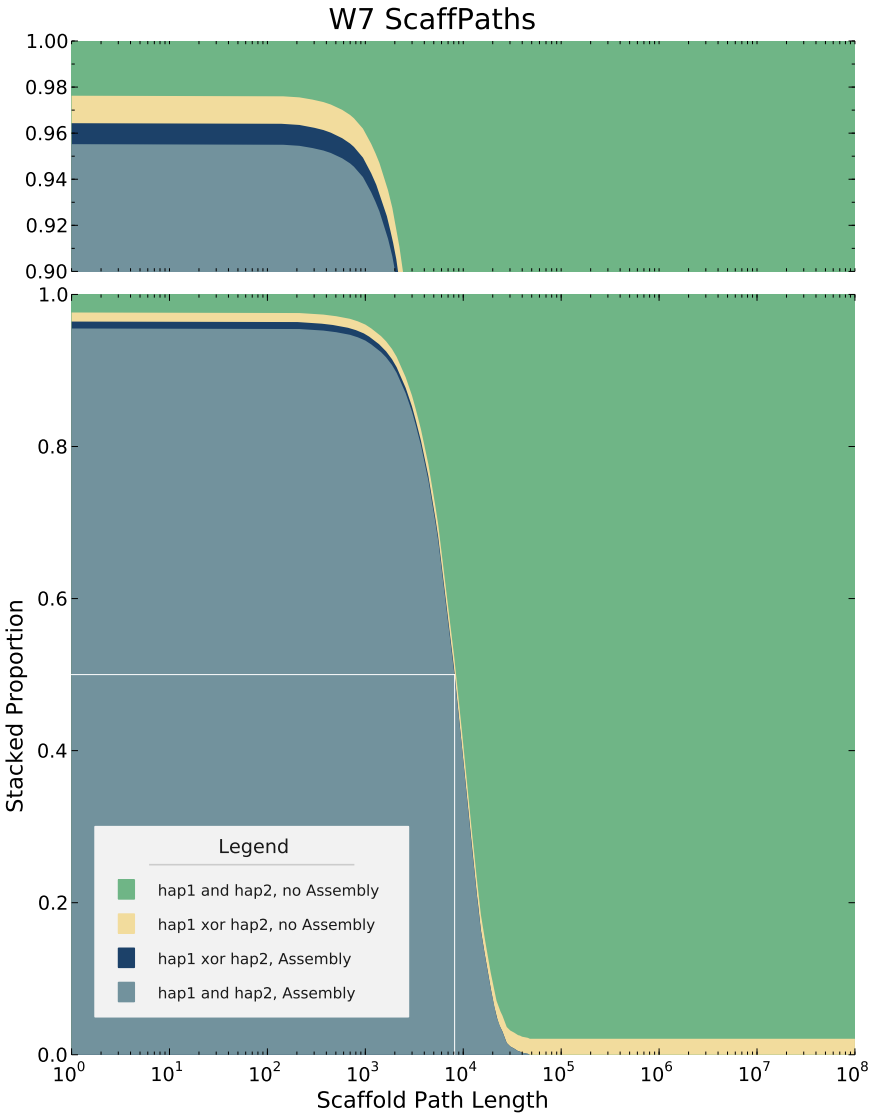


Figure 3.212: W7 scaffolds caption goes here.

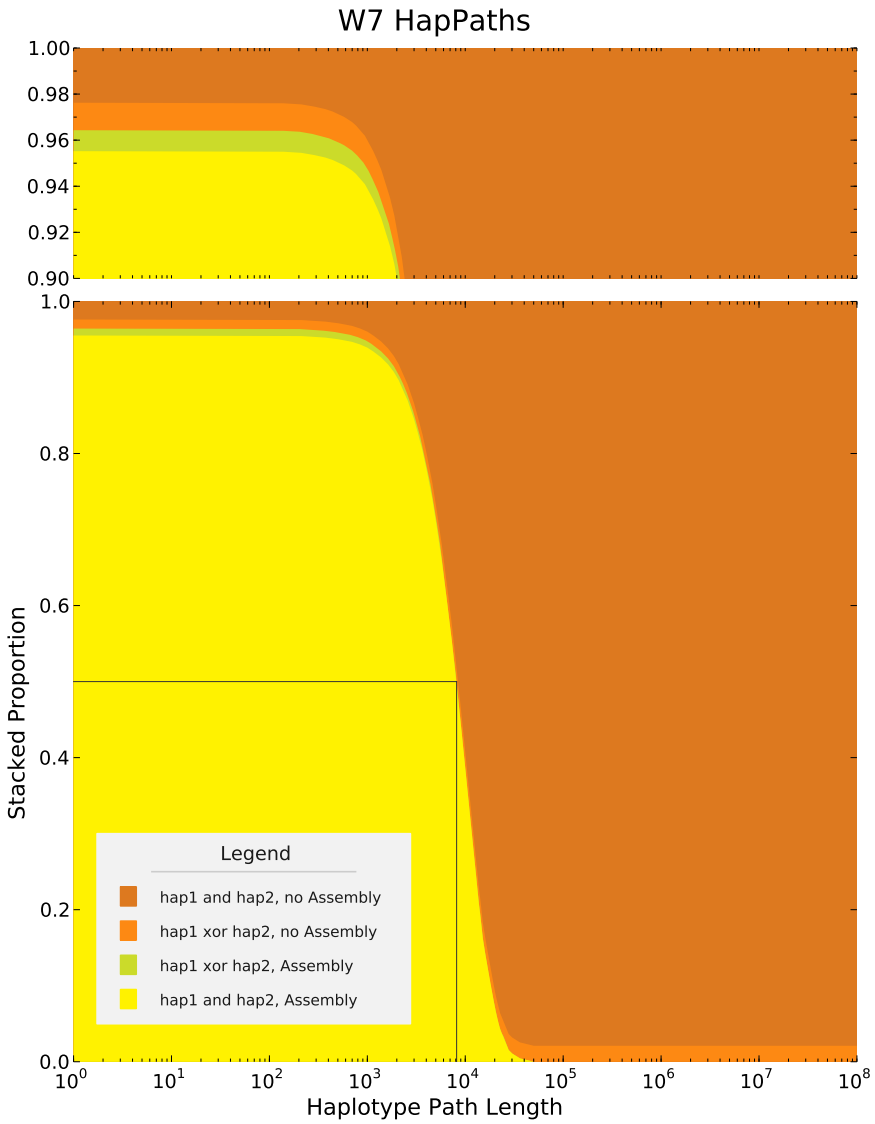


Figure 3.213: W7 hapPaths caption goes here.

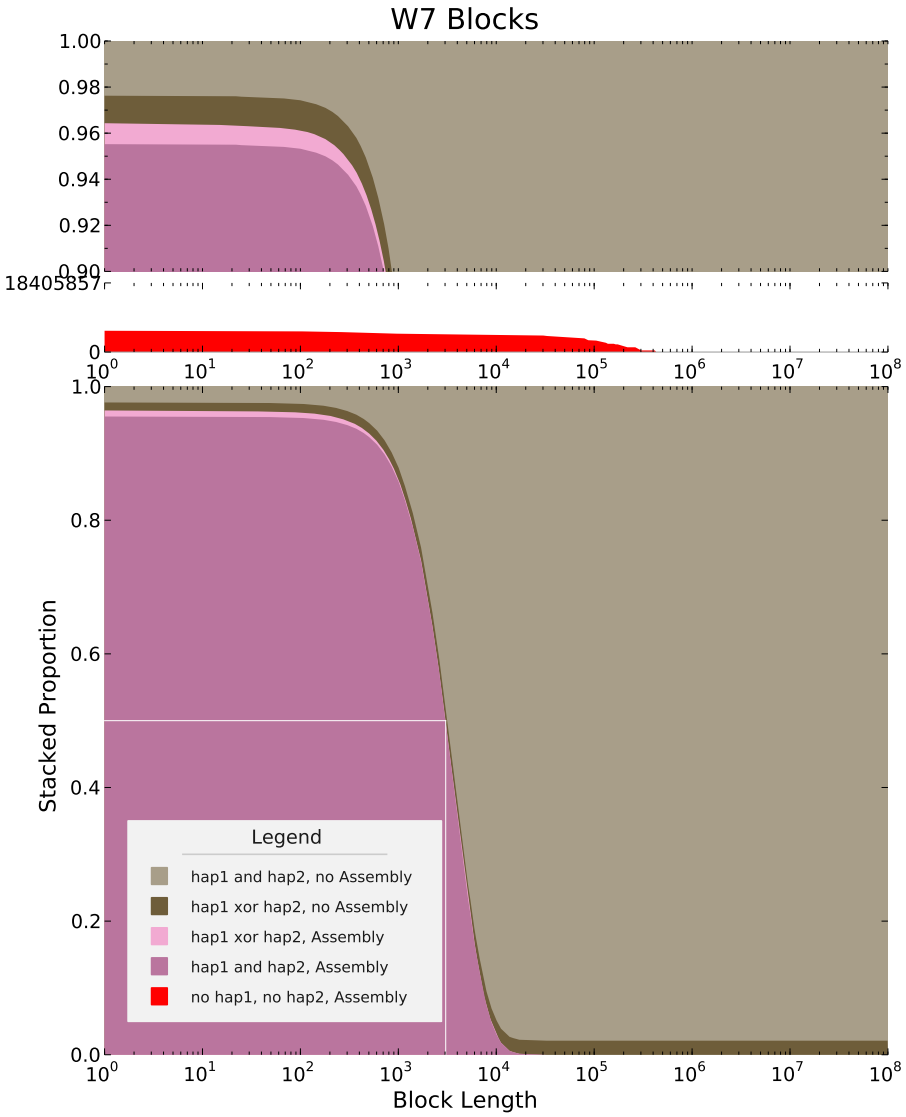


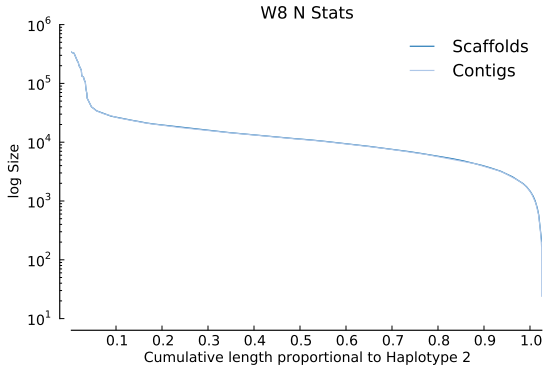
Figure 3.214: W7 blocks caption goes here.

W8

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
X1	0.97305	0.97332	0.97277	0.99869
W8	0.97204	0.97232	0.97175	0.99831
W11	0.97203	0.97220	0.97187	0.99798

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	18,213	200	1,550.00	4,251	6,337.47	8,809.00	339,806	9,147.95	115,424,301
Contigs	18,725	24	1,394.00	4,077	6,163.32	8,622.00	339,806	9,052.91	115,408,235

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,193,708 – 110,234,437	107,139,516 – 107,848,948	214,252,932.0 – 215,653,668.0	1,583 – 4,262
Heterozygous	422,545 – 439,958	412,279 – 419,949	824,492.0 – 839,722.0	9 – 39
Indel	2,044,907 – 2,420,601	882,640 – 1,029,262	1,762,628.0 – 2,055,006.0	1,226 – 1,451

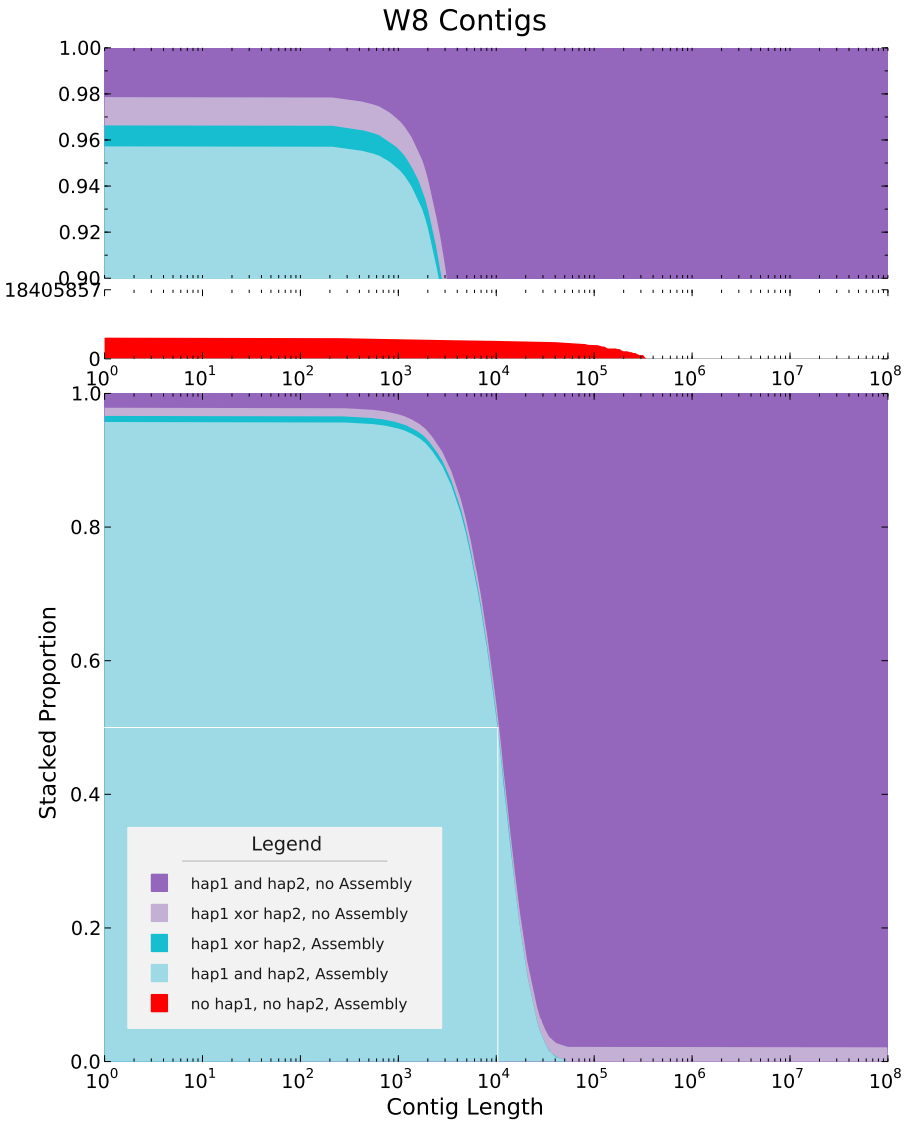


Figure 3.215: W8 contigs caption goes here.

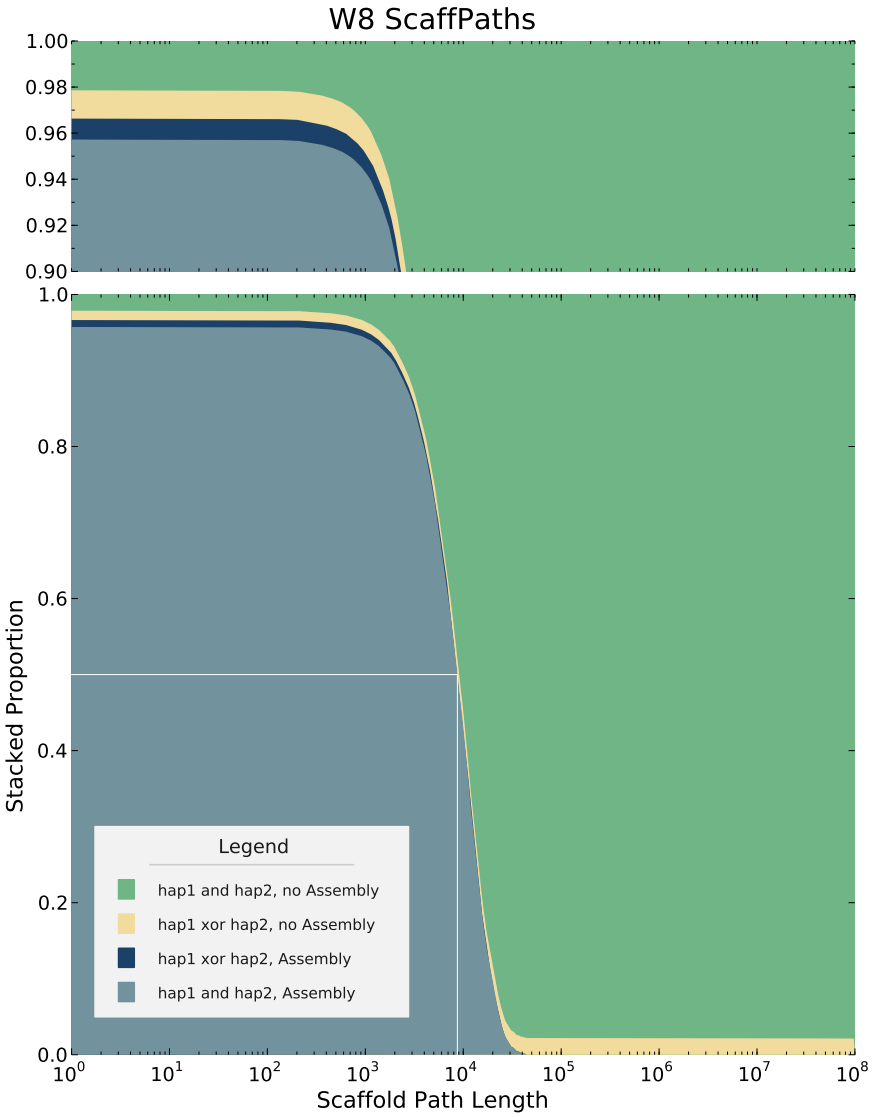


Figure 3.216: W8 scaffolds caption goes here.

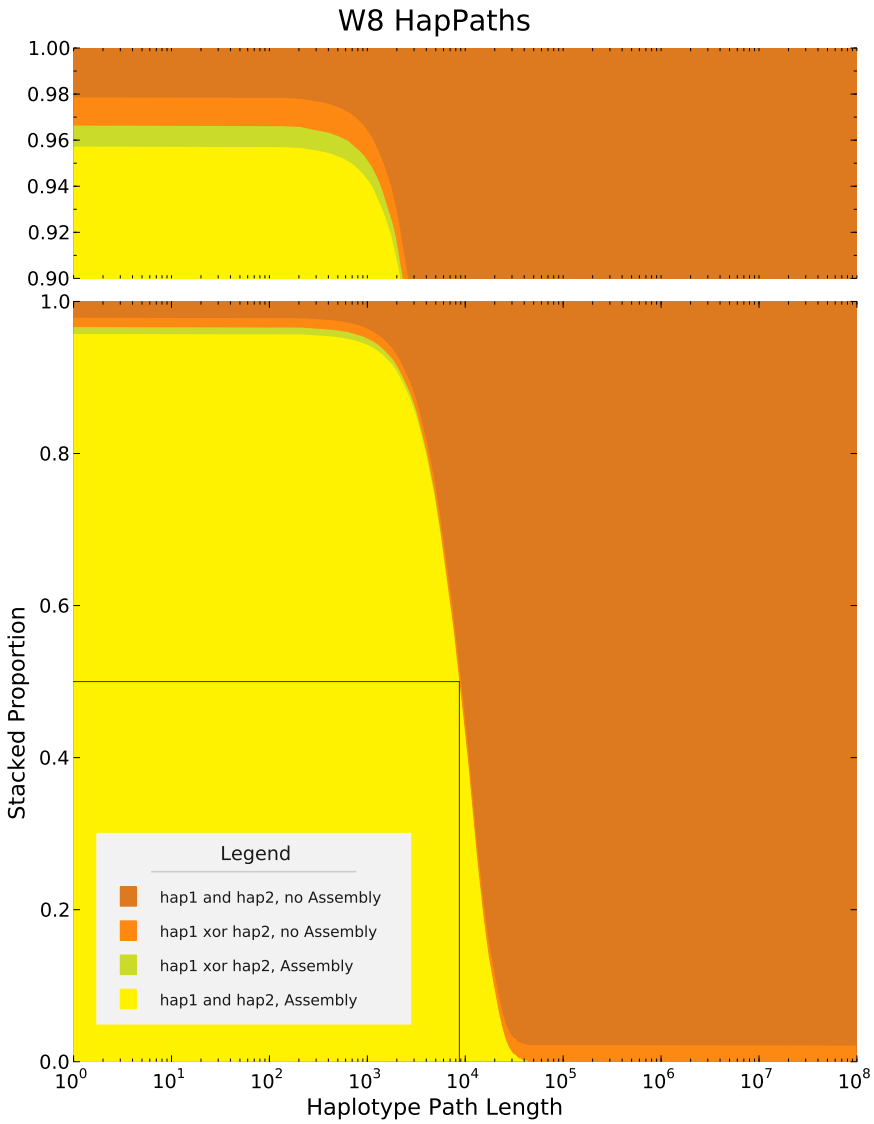


Figure 3.217: W8 hapPaths caption goes here.

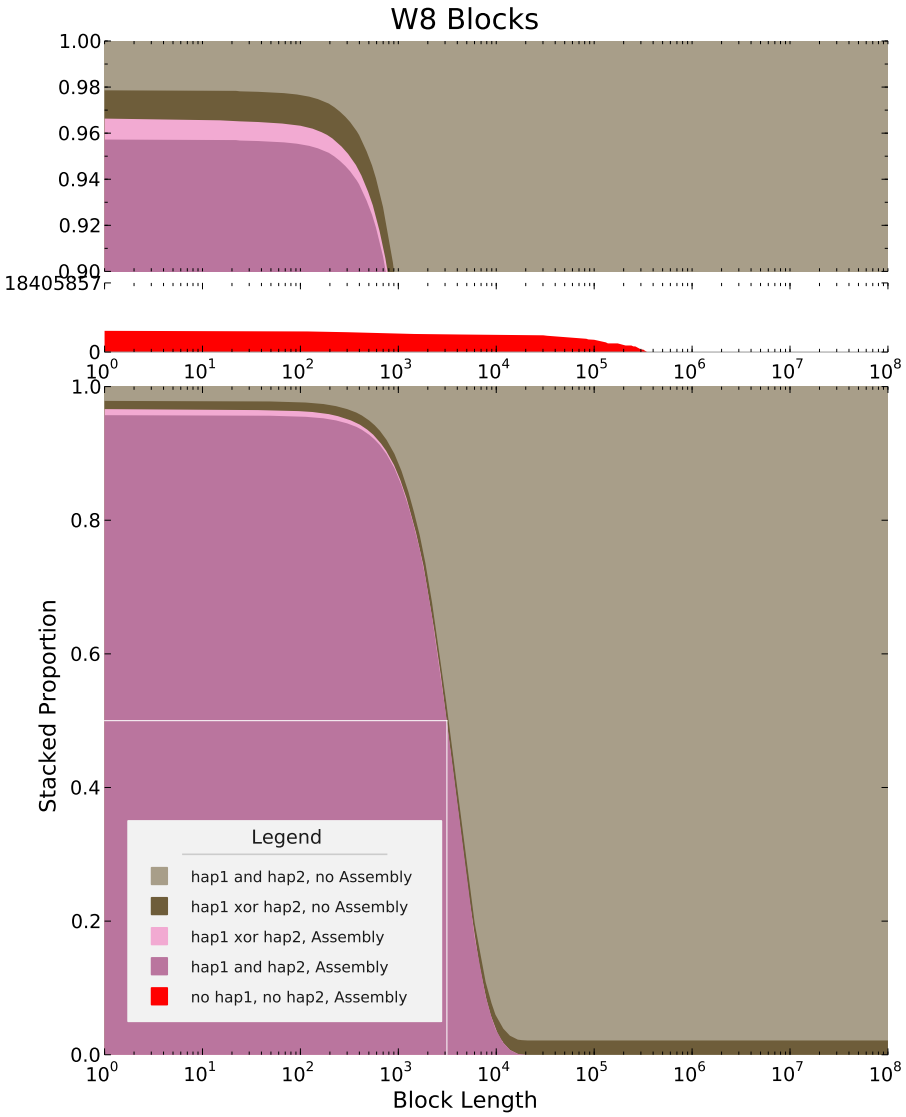


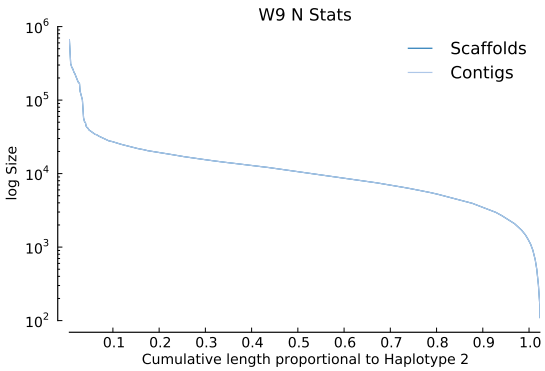
Figure 3.218: W8 blocks caption goes here.

W9

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
W5	0.97126	0.97152	0.97102	0.99800
W9	0.97080	0.97102	0.97057	0.99805
W1	0.97034	0.97048	0.97023	0.99825

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	19,862	110	1,303.00	3,680	5,797.73	7,888.50	665,681	9,471.16	115,154,477
Contigs	19,862	110	1,303.00	3,680	5,797.73	7,888.50	665,681	9,471.16	115,154,477

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,213,229 – 110,232,757	107,044,822 – 107,701,039	214,084,260.0 – 215,391,840.0	2,692 – 5,119
Heterozygous	422,097 – 439,967	411,322 – 417,889	822,614.0 – 835,722.0	15 – 28
Indel	1,955,533 – 2,328,994	841,107 – 969,162	1,679,592.0 – 1,935,354.0	1,311 – 1,485

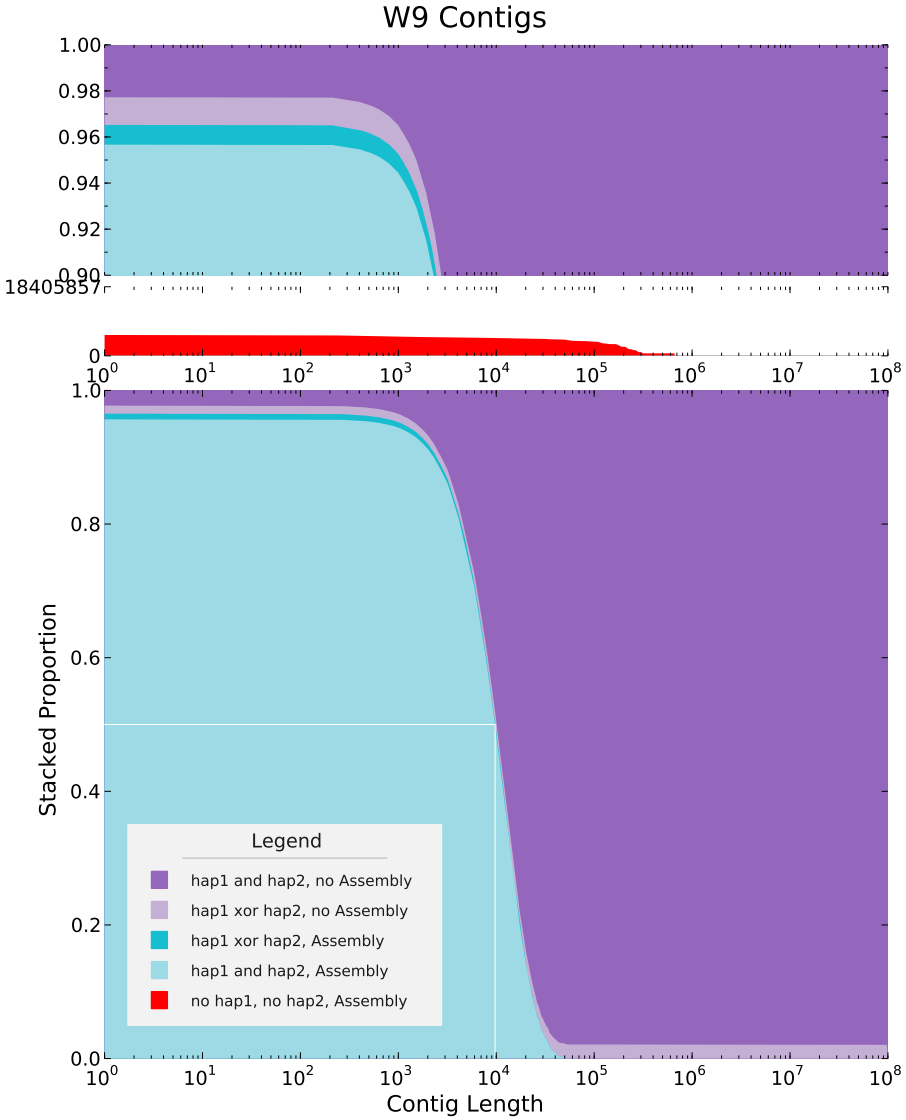


Figure 3.219: W9 contigs caption goes here.

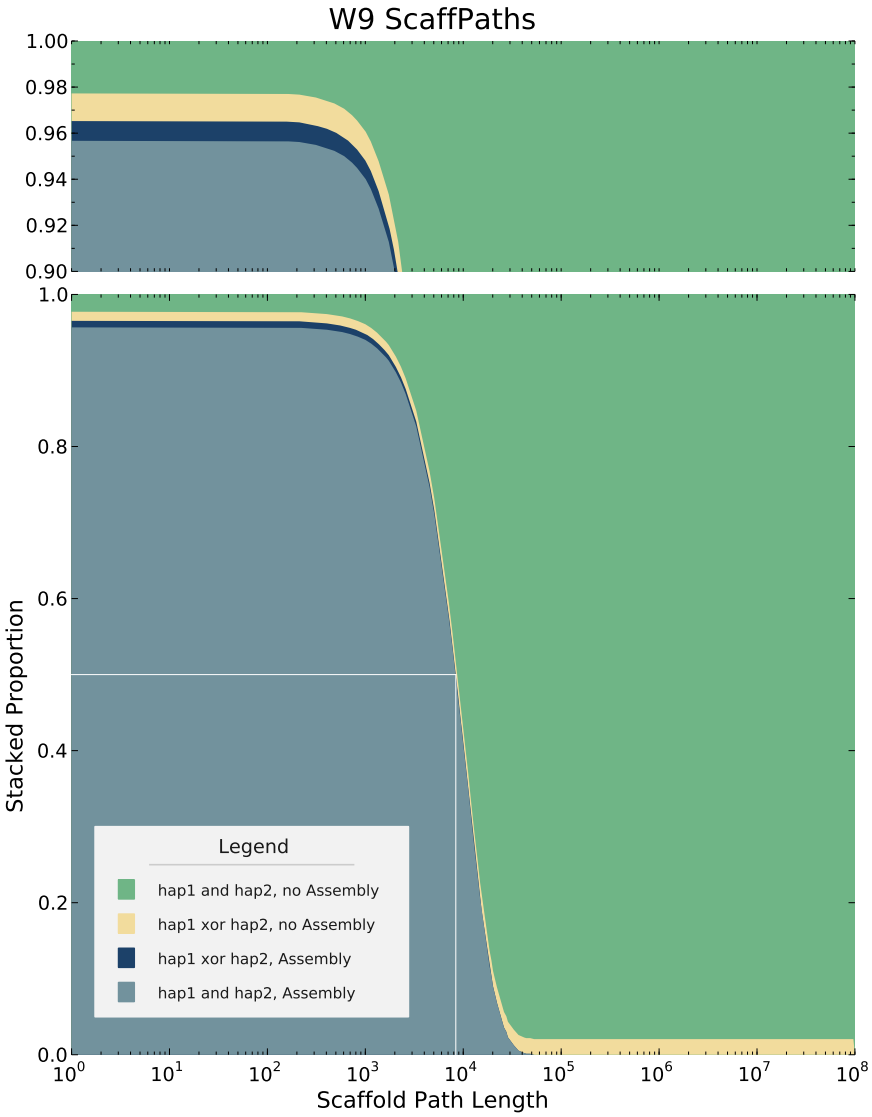


Figure 3.220: W9 scaffolds caption goes here.

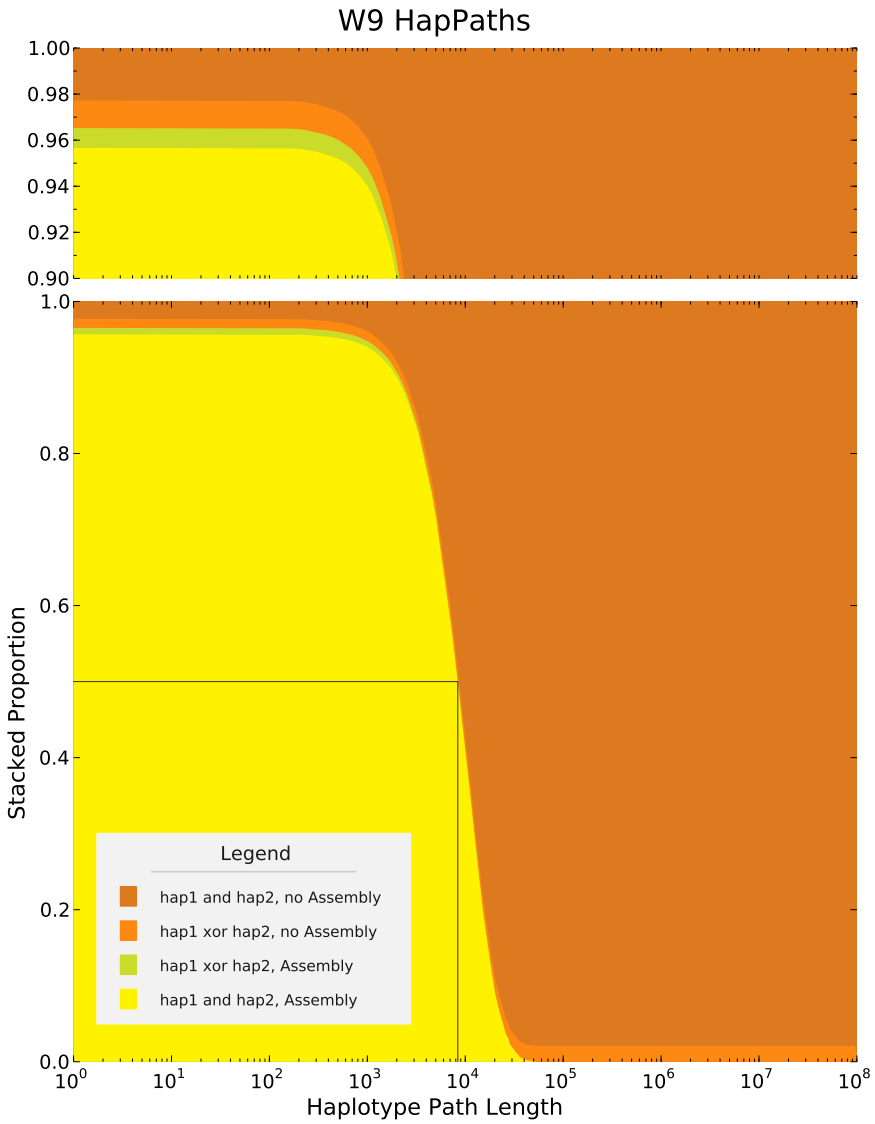


Figure 3.221: W9 hapPaths caption goes here.

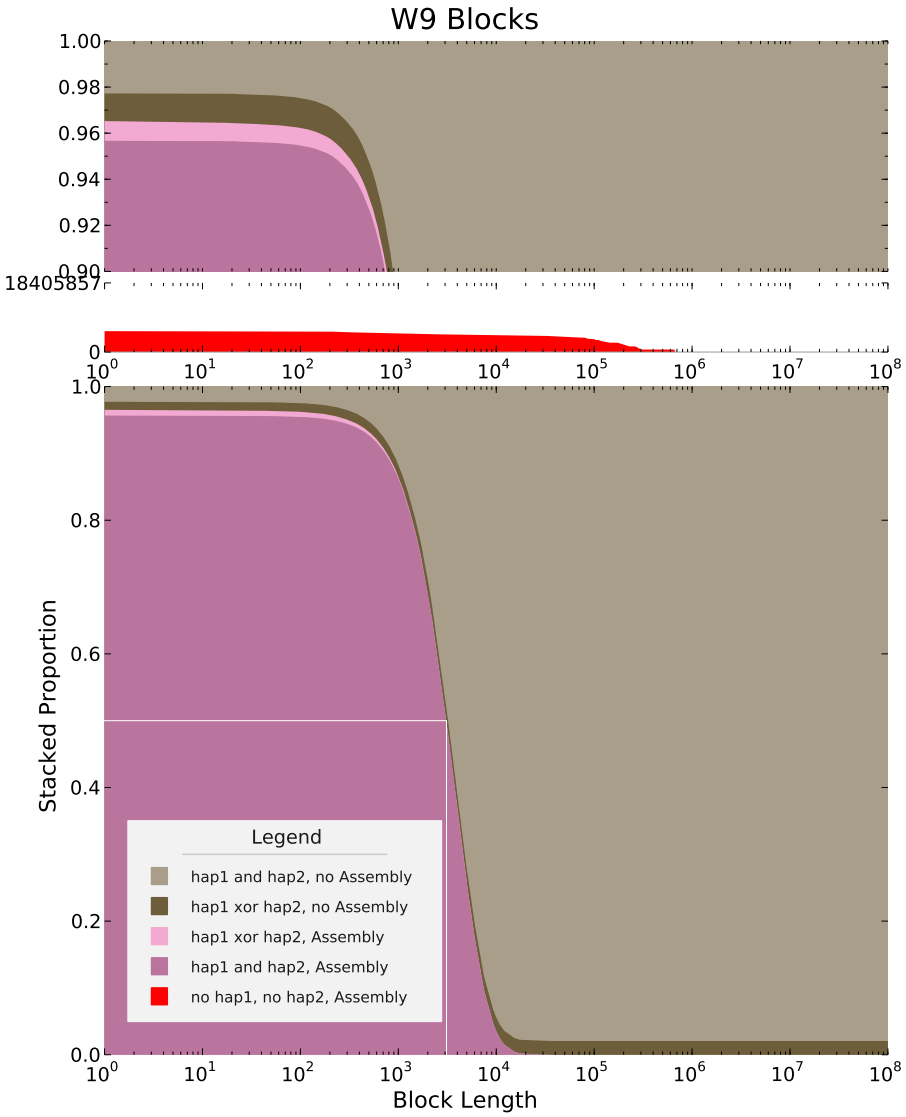


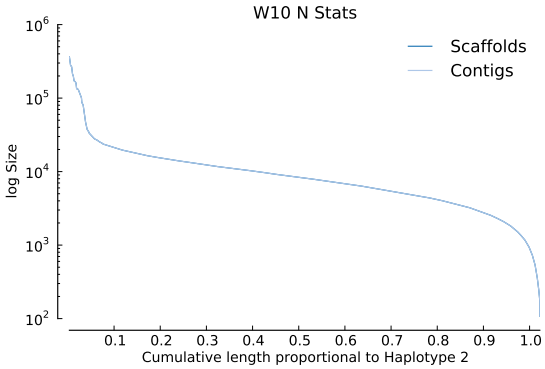
Figure 3.222: W9 blocks caption goes here.

W10

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
W6	0.96892	0.96918	0.96865	0.99764
W10	0.96812	0.96852	0.96772	0.99812
X5	0.96766	0.96789	0.96742	0.99789

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	24,778	108	1,072.00	2,992	4,642.18	6,269.75	364,958	7,006.89	115,023,920
Contigs	24,778	108	1,072.00	2,992	4,642.18	6,269.75	364,958	7,006.89	115,023,920

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,152,890 – 110,261,108	106,751,246 – 107,445,597	213,497,746.0 – 214,880,856.0	2,373 – 5,169
Heterozygous	420,179 – 439,680	408,350 – 415,325	816,660.0 – 830,562.0	20 – 44
Indel	1,852,801 – 2,224,890	774,357 – 898,074	1,546,144.0 – 1,793,208.0	1,285 – 1,470

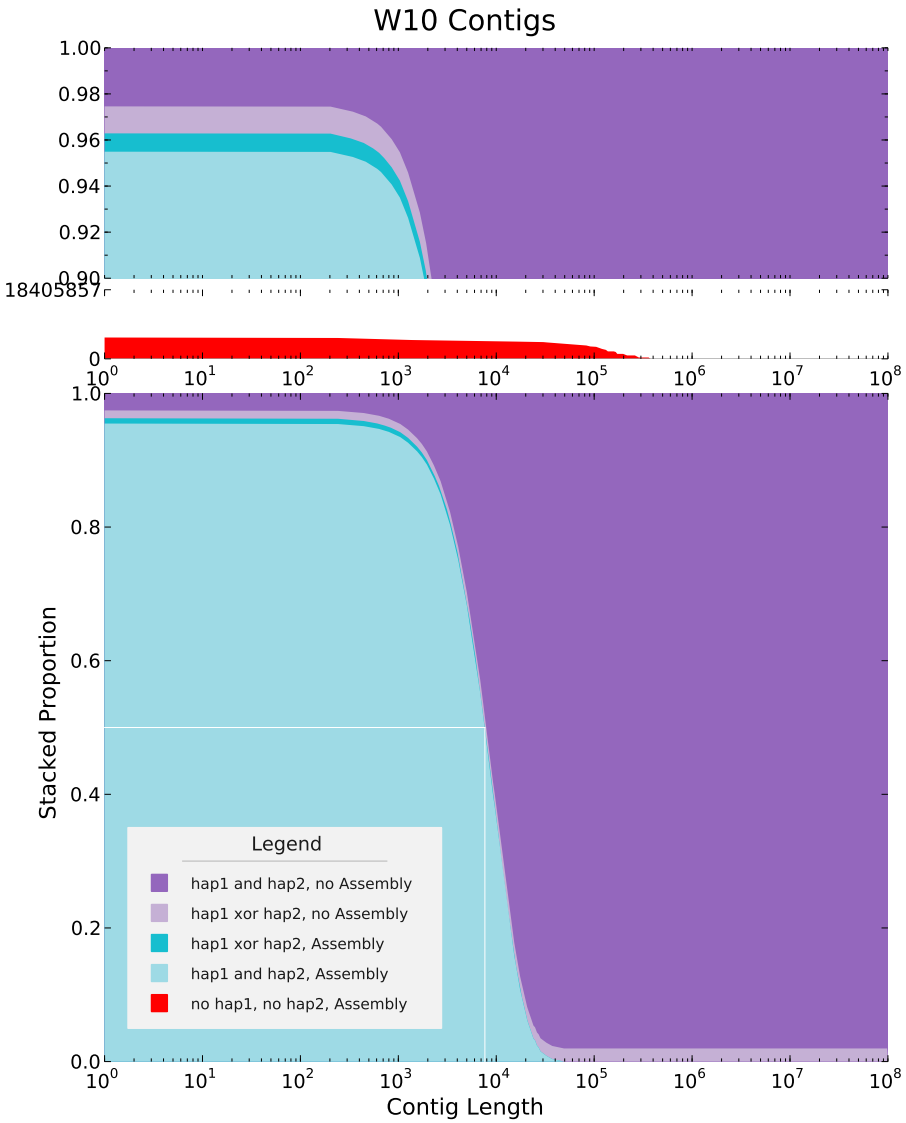


Figure 3.223: W10 contigs caption goes here.

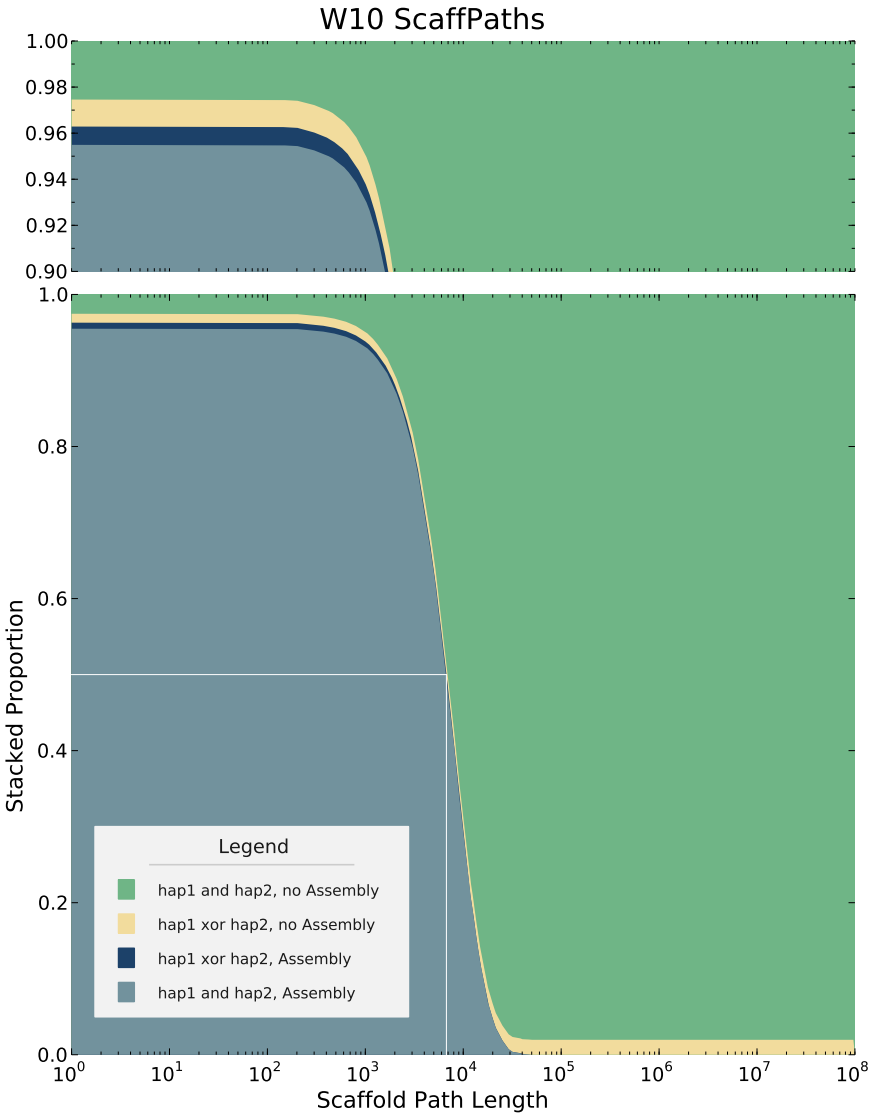


Figure 3.224: W10 scaffolds caption goes here.

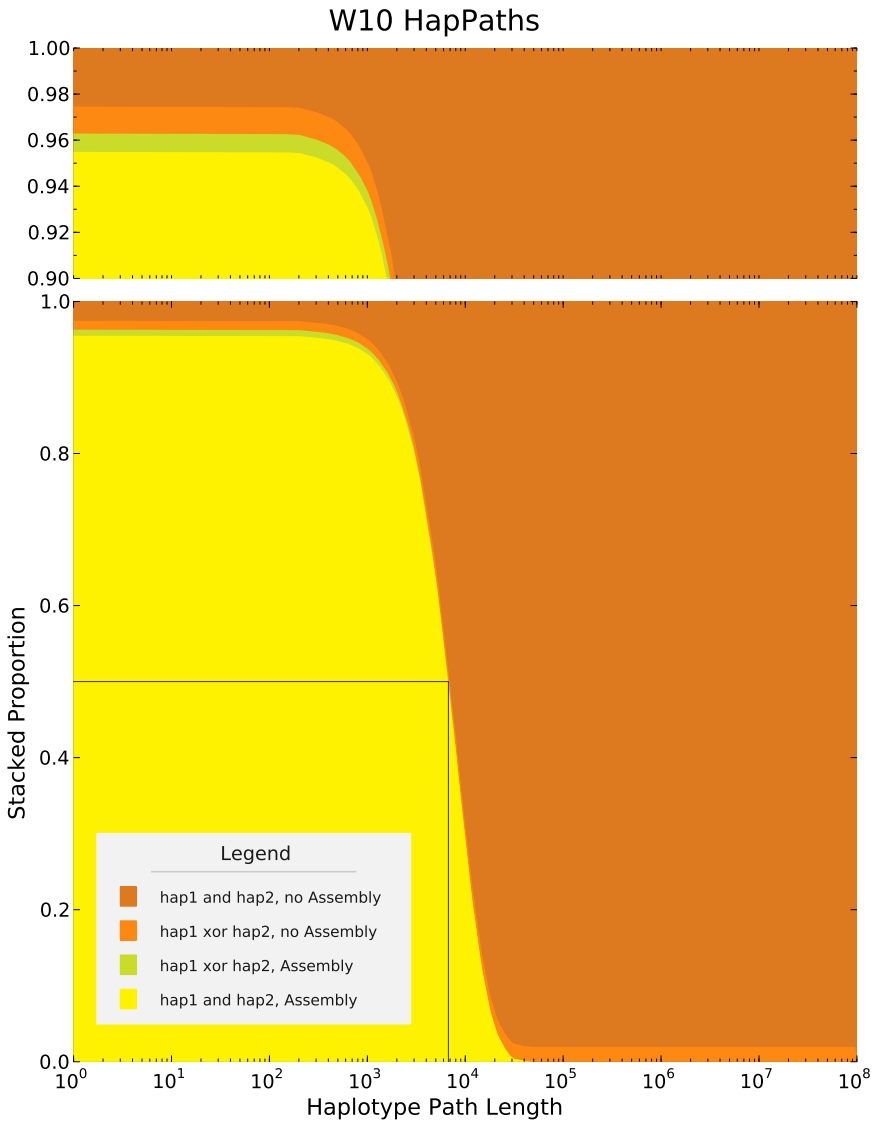


Figure 3.225: W10 hapPaths caption goes here.

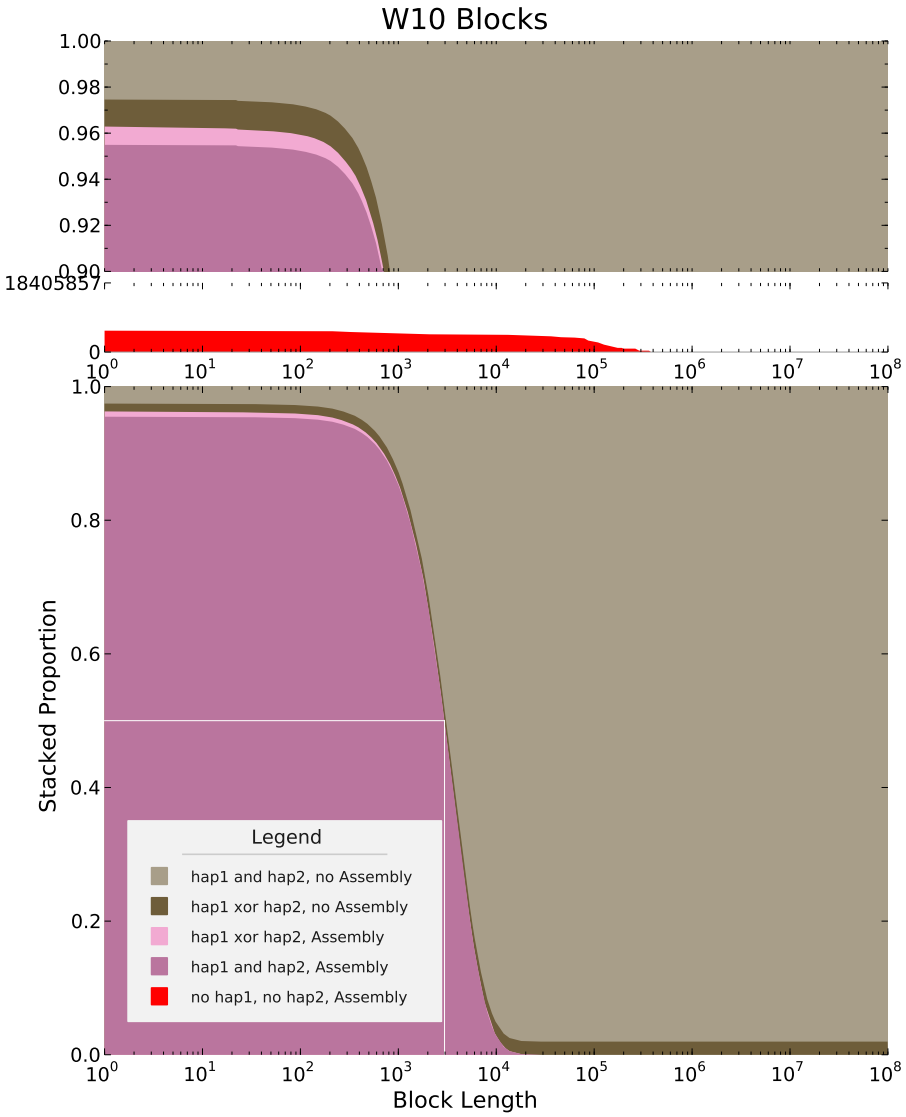


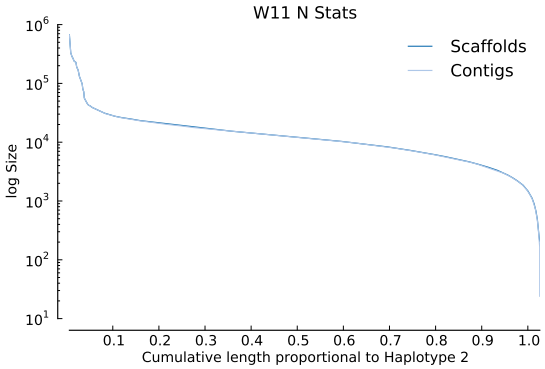
Figure 3.226: W10 blocks caption goes here.

W11

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
W8	0.97204	0.97232	0.97175	0.99831
W11	0.97203	0.97220	0.97187	0.99798
W5	0.97126	0.97152	0.97102	0.99800

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	17,581	200	1,450.00	4,139	6,566.96	9,355.00	665,472	10,231.14	115,453,668
Contigs	17,979	24	1,334.00	4,023	6,420.90	9,178.00	665,472	10,133.33	115,441,384

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	109,194,772 – 110,214,561	107,129,828 – 107,831,708	214,235,328.0 – 215,619,862.0	1,616 – 4,246
Heterozygous	422,826 – 439,880	412,509 – 420,118	824,954.0 – 840,056.0	9 – 38
Indel	2,069,244 – 2,446,102	899,019 – 1,046,181	1,795,430.0 – 2,088,912.0	1,217 – 1,423

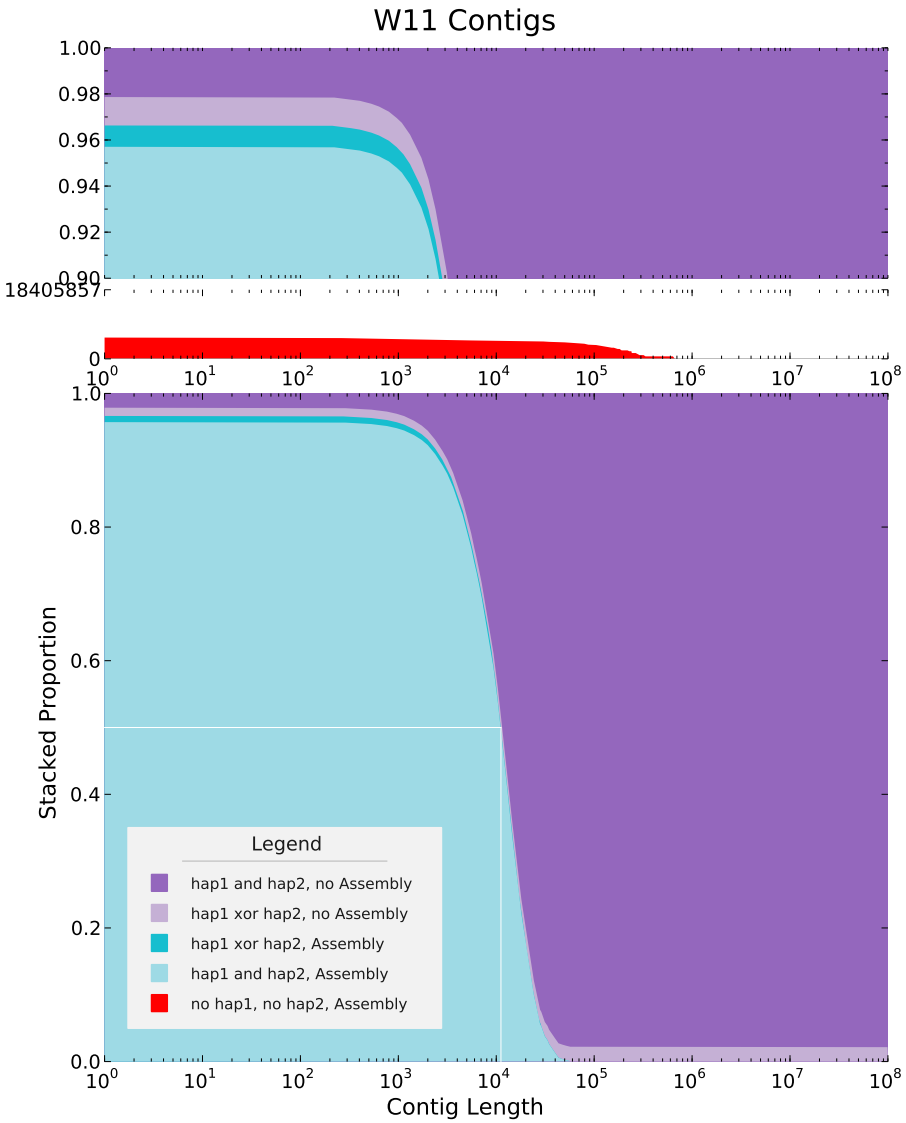


Figure 3.227: W11 contigs caption goes here.

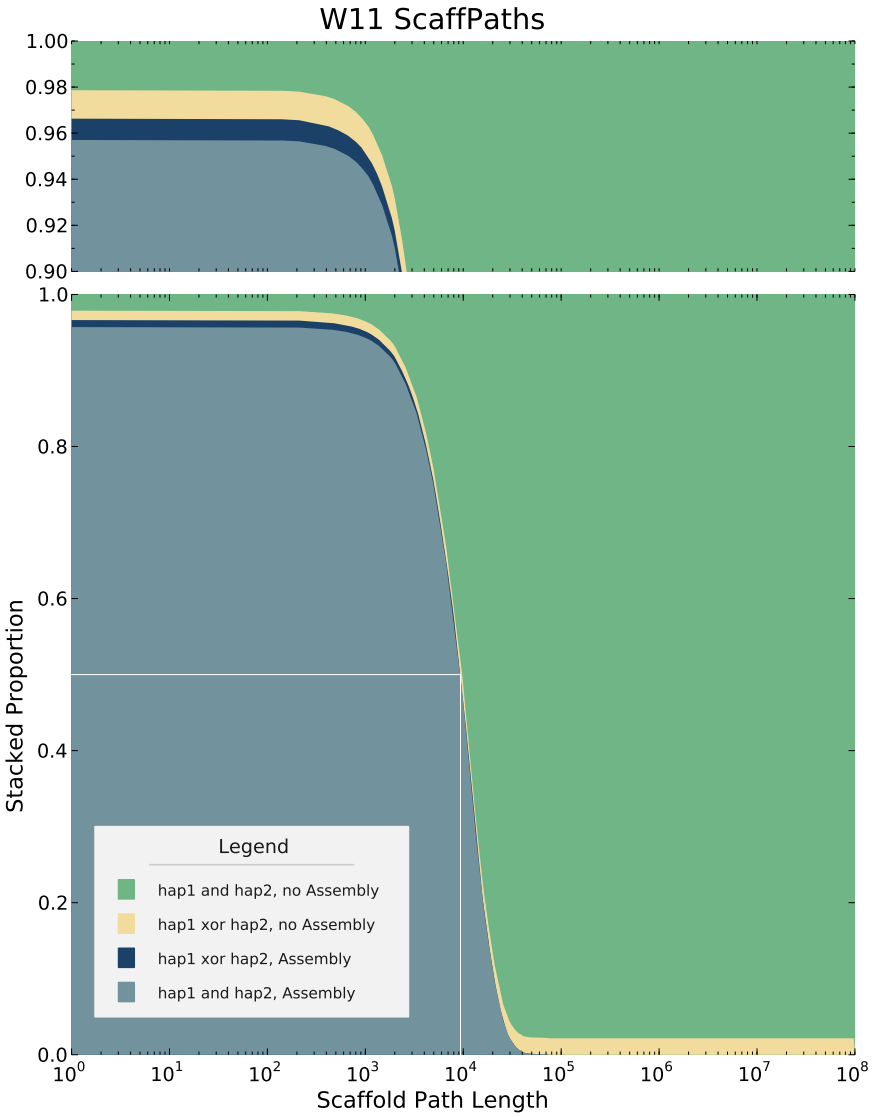


Figure 3.228: W11 scaffolds caption goes here.

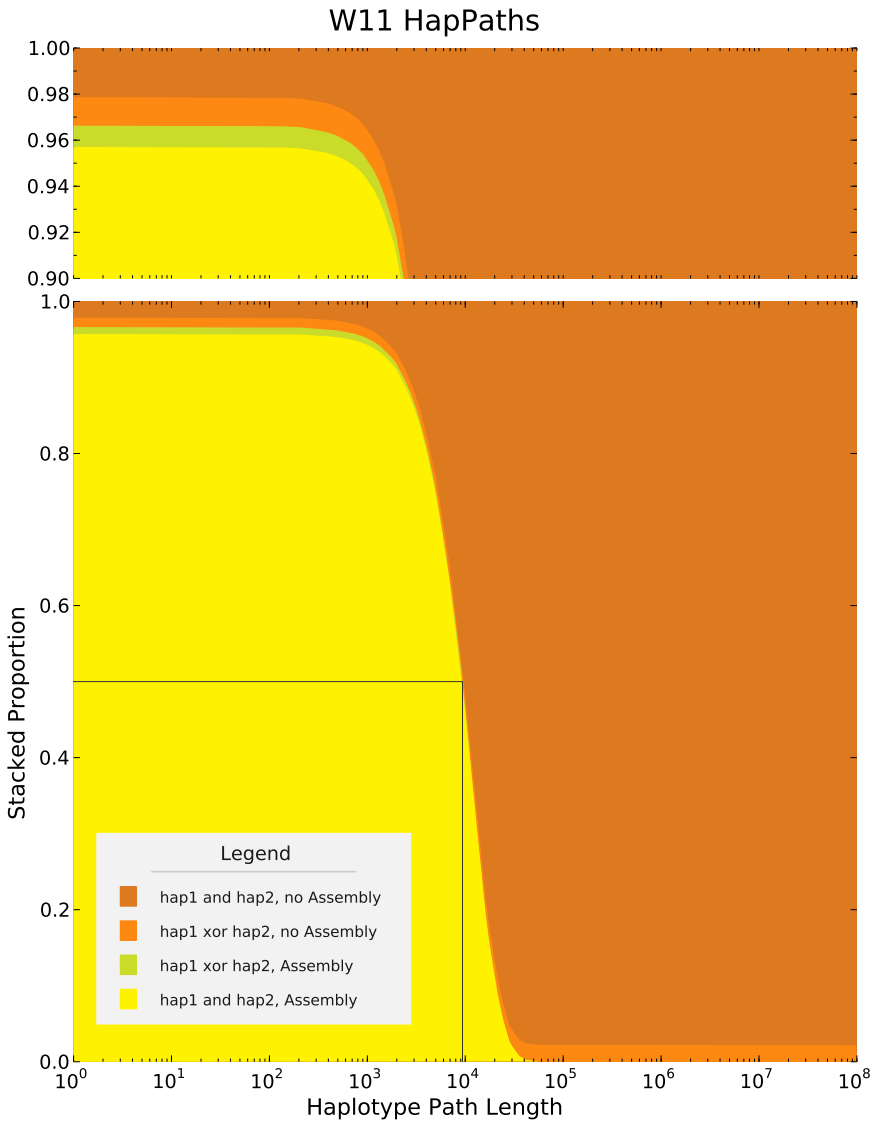


Figure 3.229: W11 hapPaths caption goes here.

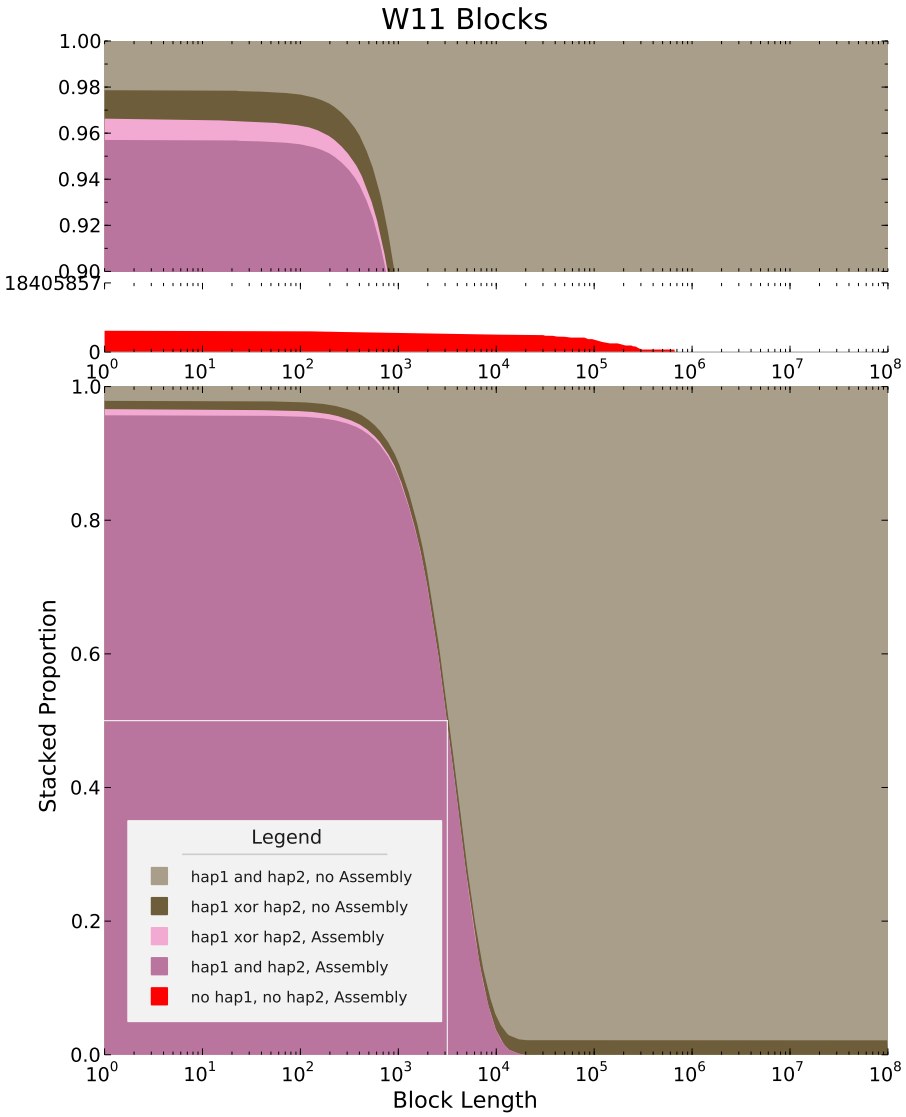


Figure 3.230: W11 blocks caption goes here.

3.2.20 X, Auto

Affiliation: Auto
Contact: Auto
Software: **ABySS**
Number of entries: 6

ID	Total	Hap 1	Hap 2	Bac
X2	0.97492	0.97516	0.97467	0.99848
X1	0.97305	0.97332	0.97277	0.99869
X5	0.96766	0.96789	0.96742	0.99789
X4	0.96758	0.96779	0.96735	0.99798
X6	0.95516	0.95527	0.95504	0.99562
X3	0.95516	0.95535	0.95495	0.99560

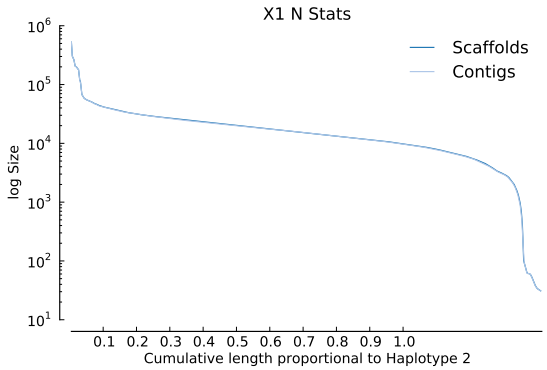
Assemblies:

X1

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
C1	0.97348	0.97347	0.97348	0.00903
X1	0.97305	0.97332	0.97277	0.99869
W8	0.97204	0.97232	0.97175	0.99831

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	151,203	31	33.00	42	1,054.09	61.00	524,383	4,782.20	159,381,550
Contigs	151,852	31	33.00	43	1,049.29	61.00	524,383	4,756.40	159,336,958

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	80,406,978 – 81,196,668	78,360,674 – 78,873,243	156,719,343.7 – 157,743,082.7	672 – 907
Heterozygous	309,868 – 324,469	299,589 – 305,889	597,192.2 – 609,240.2	947 – 1,190
Indel	1,801,359 – 2,129,752	654,492 – 864,407	1,307,197.0 – 1,725,831.0	846 – 926

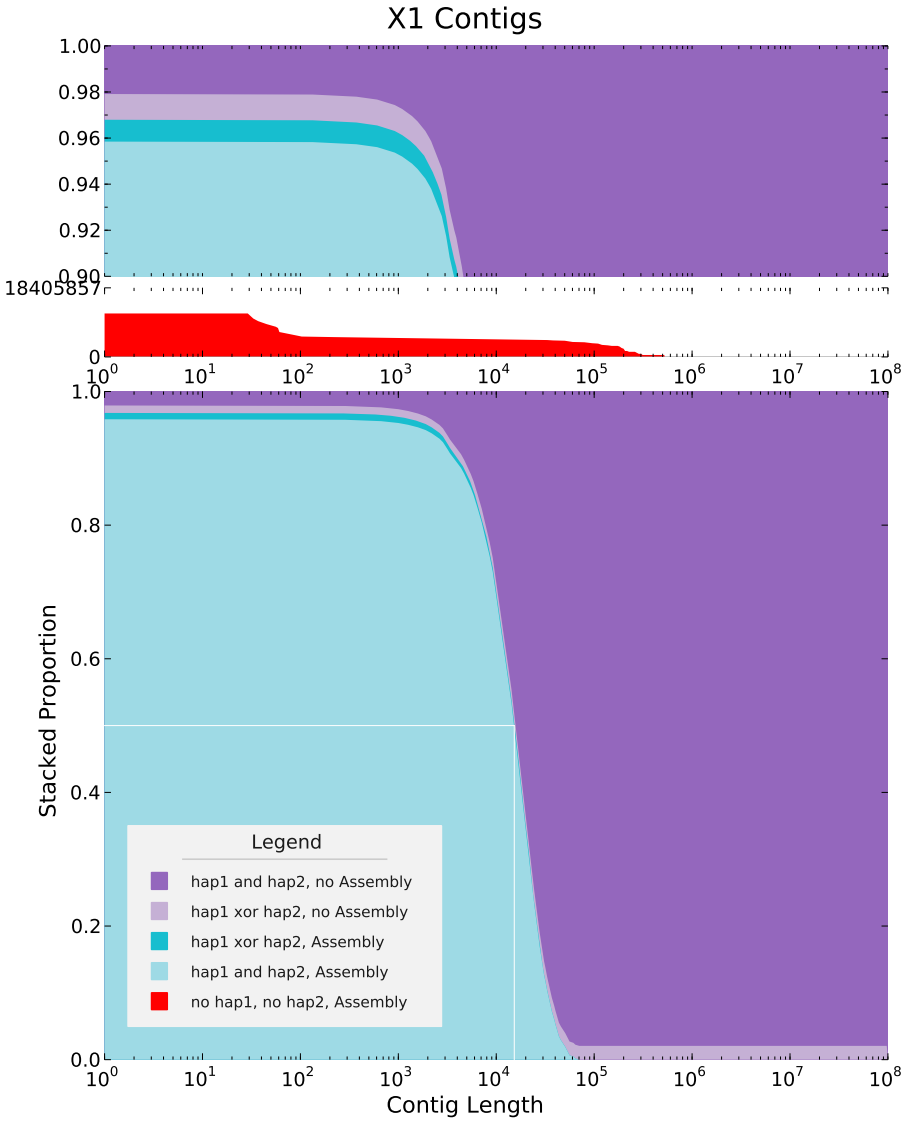


Figure 3.231: X1 contigs caption goes here.

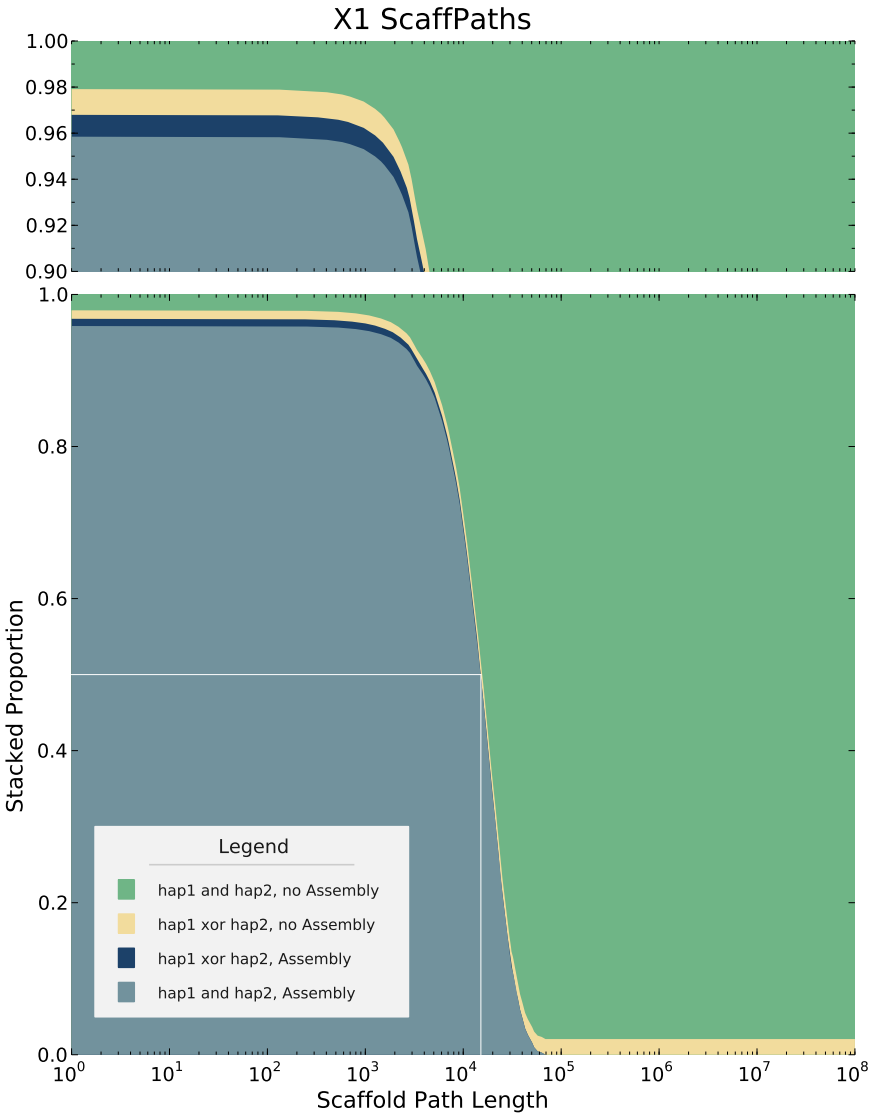


Figure 3.232: X1 scaffolds caption goes here.

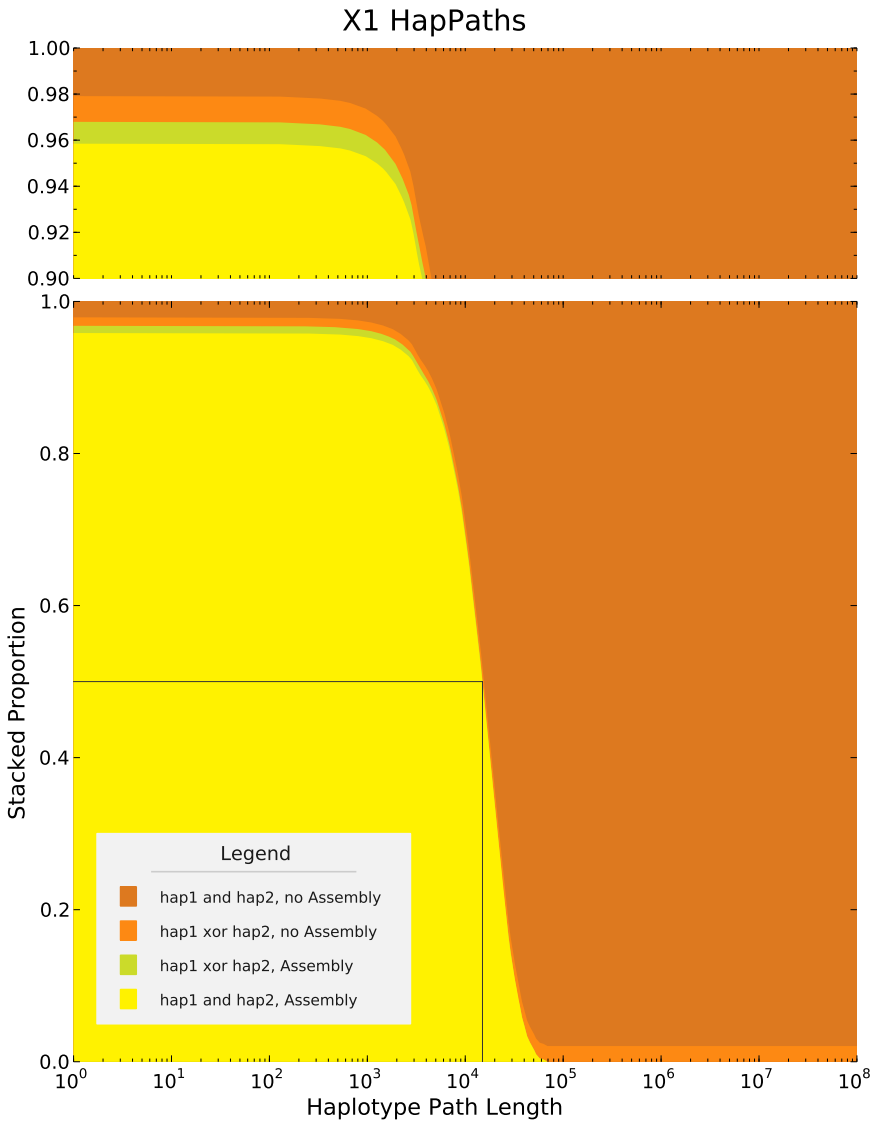


Figure 3.233: X1 hapPaths caption goes here.

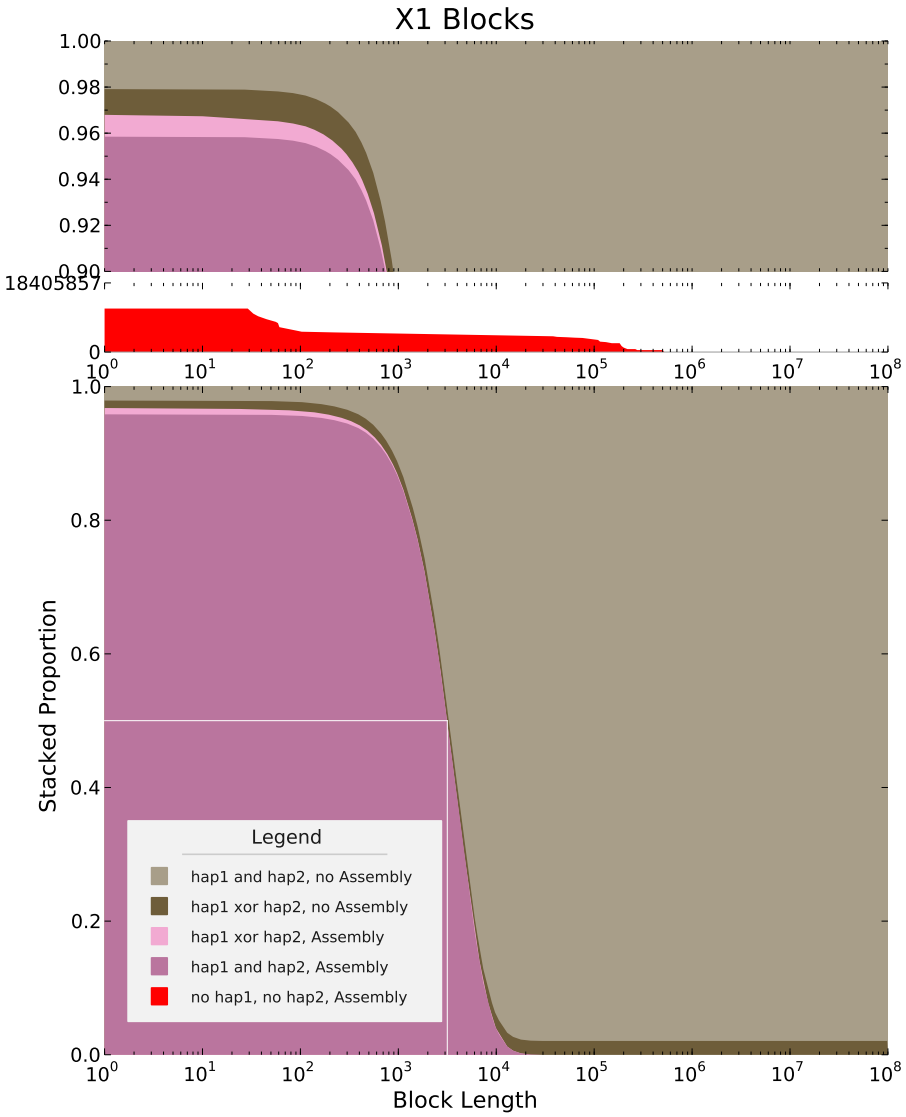


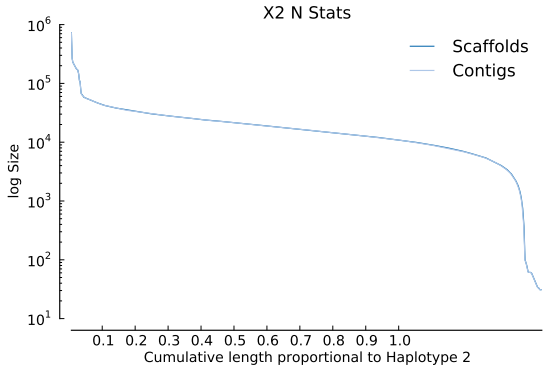
Figure 3.234: X1 blocks caption goes here.

X2

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
D3	0.97616	0.97634	0.97595	0.99085
X2	0.97492	0.97516	0.97467	0.99848
C1	0.97348	0.97347	0.97348	0.00903

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	140,598	31	33.00	43	1,148.16	61.00	722,867	5,188.27	161,428,633
Contigs	141,144	31	33.00	43	1,143.47	61.00	722,867	5,169.49	161,394,359

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	79,239,233 – 79,990,966	77,382,200 – 77,871,325	154,762,522.8 – 155,739,569.2	628 – 822
Heterozygous	305,272 – 319,306	295,941 – 302,179	589,799.9 – 601,727.9	994 – 1,240
Indel	1,840,656 – 2,166,723	700,225 – 909,364	1,398,424.0 – 1,815,350.4	968 – 1,049

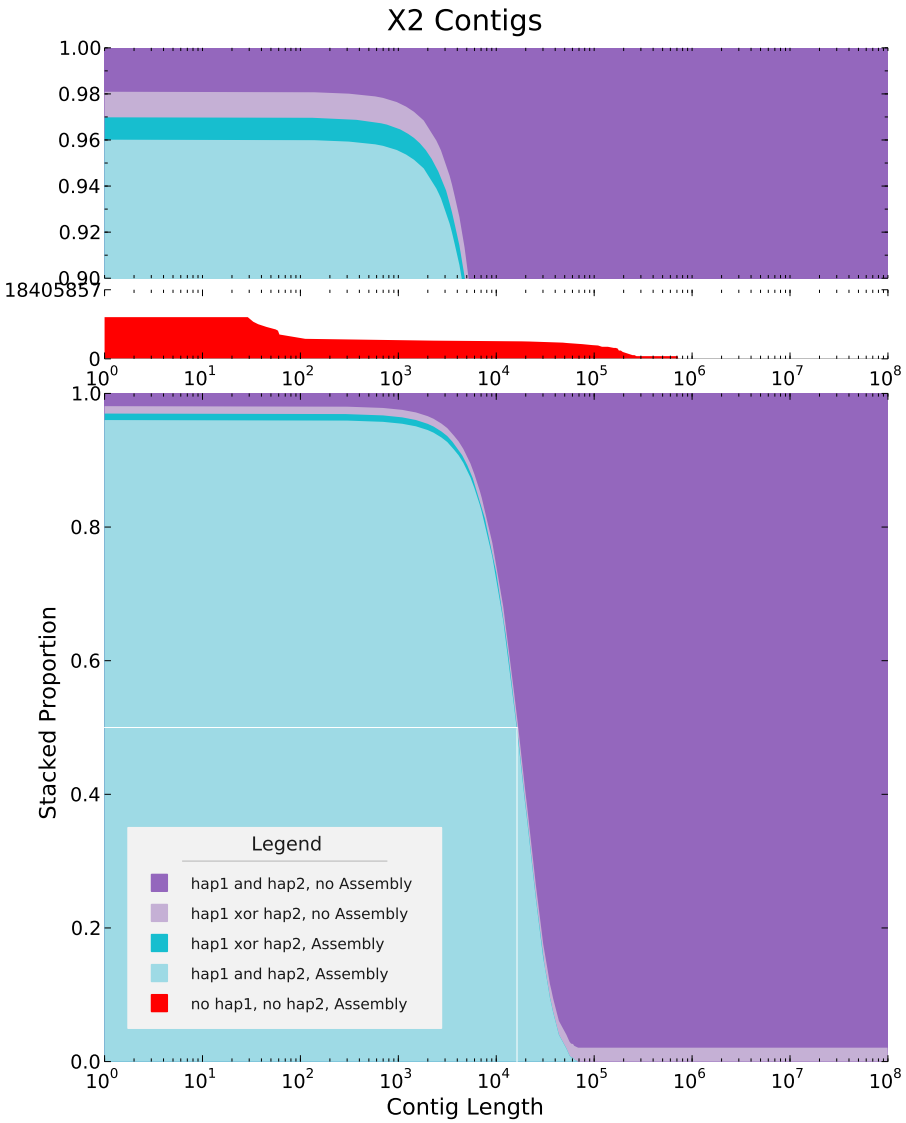


Figure 3.235: X2 contigs caption goes here.

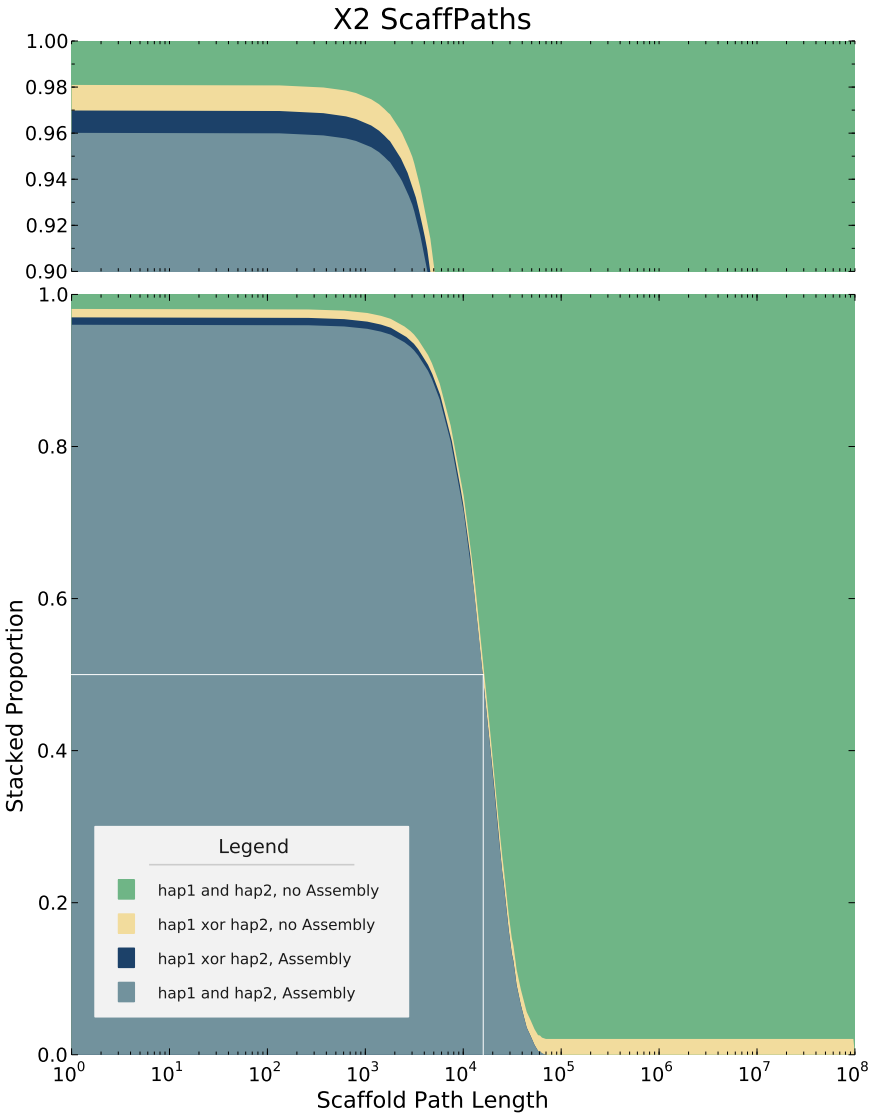


Figure 3.236: X2 scaffolds caption goes here.

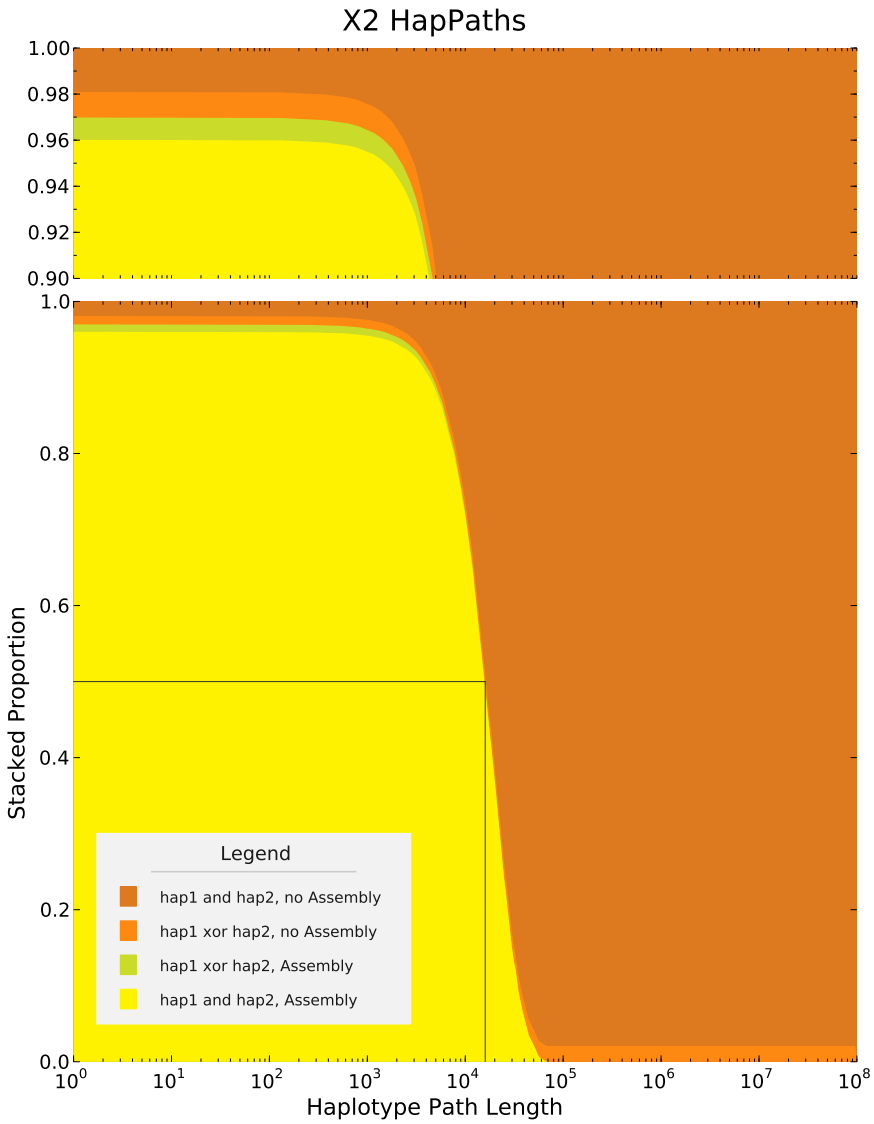


Figure 3.237: X2 hapPaths caption goes here.

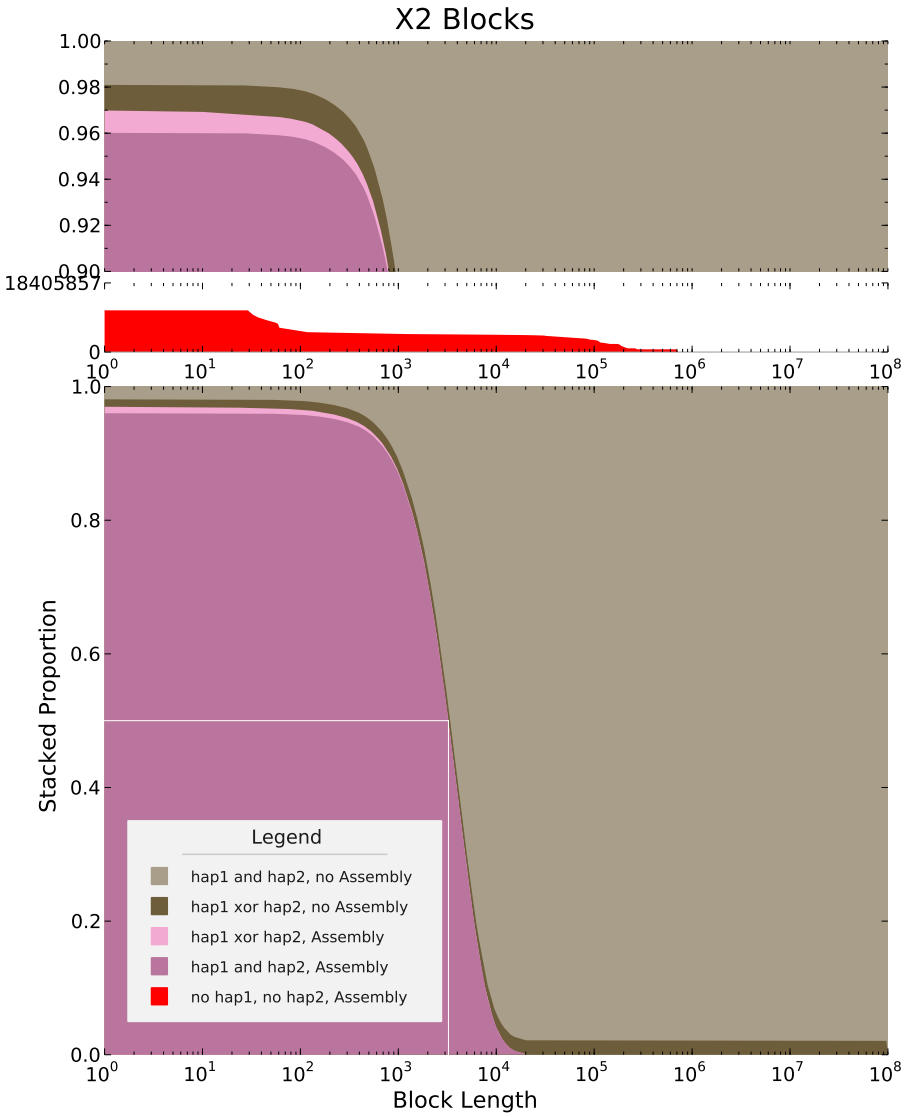


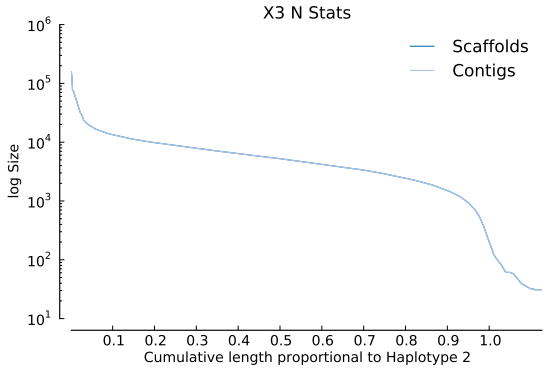
Figure 3.238: X2 blocks caption goes here.

X3

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
X6	0.95516	0.95527	0.95504	0.99562
X3	0.95516	0.95535	0.95495	0.99560
H5	0.94518	0.94532	0.94503	0.99789

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	328,797	31	32.00	39	385.31	62.00	156,082	1,565.50	126,688,003
Contigs	328,797	31	32.00	39	385.31	62.00	156,082	1,565.50	126,688,003

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,463,072 – 110,237,427	104,701,577 – 105,757,697	209,403,132.0 – 211,515,286.0	11 – 54
Heterozygous	412,504 – 436,689	394,809 – 401,023	789,618.0 – 802,046.0	0 – 0
Indel	1,158,499 – 1,529,620	443,668 – 688,821	886,156.0 – 1,376,348.0	590 – 647

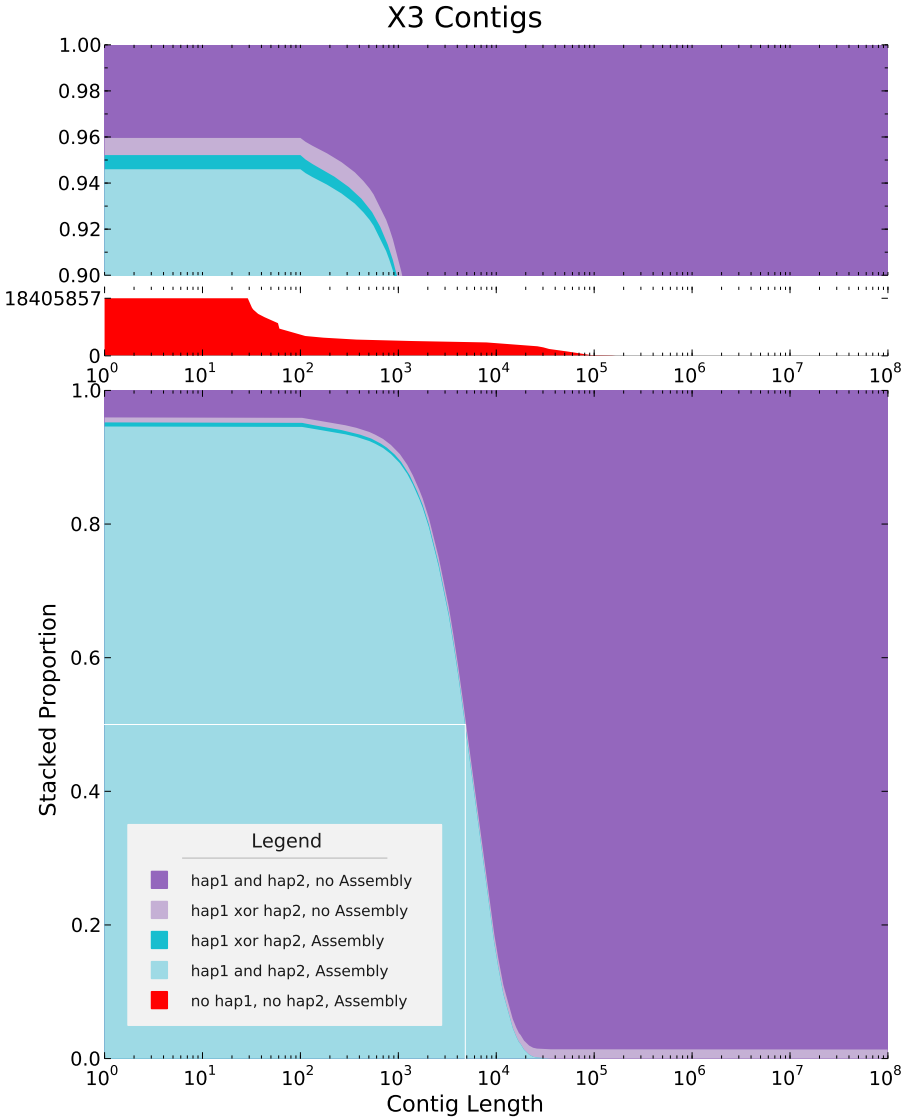


Figure 3.239: X3 contigs caption goes here.

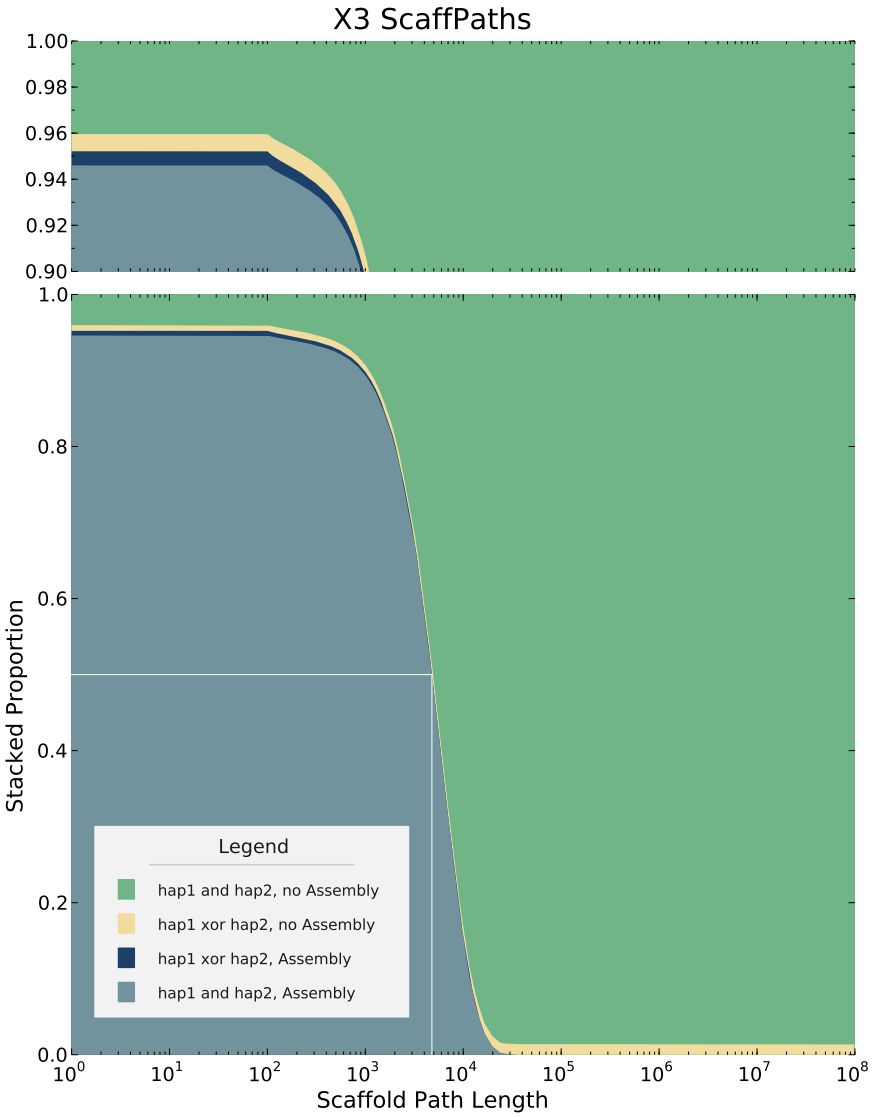


Figure 3.240: X3 scaffolds caption goes here.

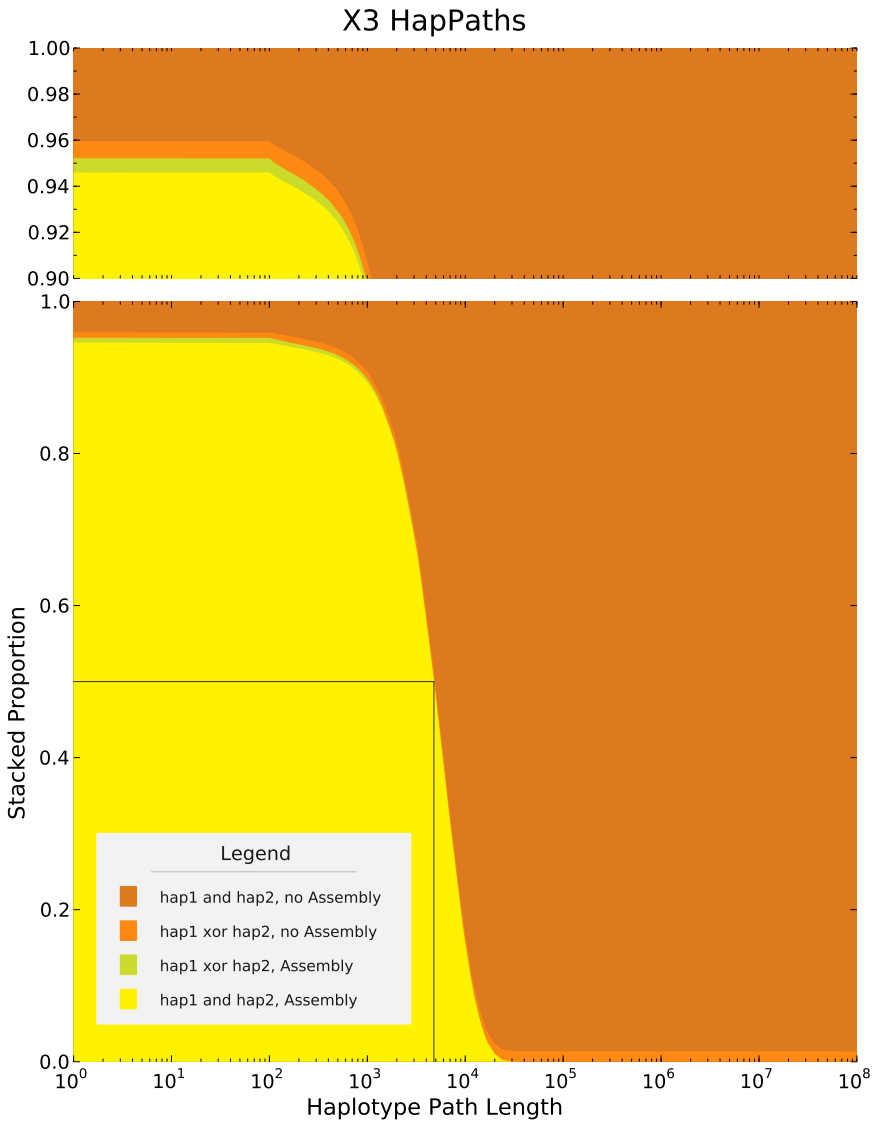


Figure 3.241: X3 hapPaths caption goes here.

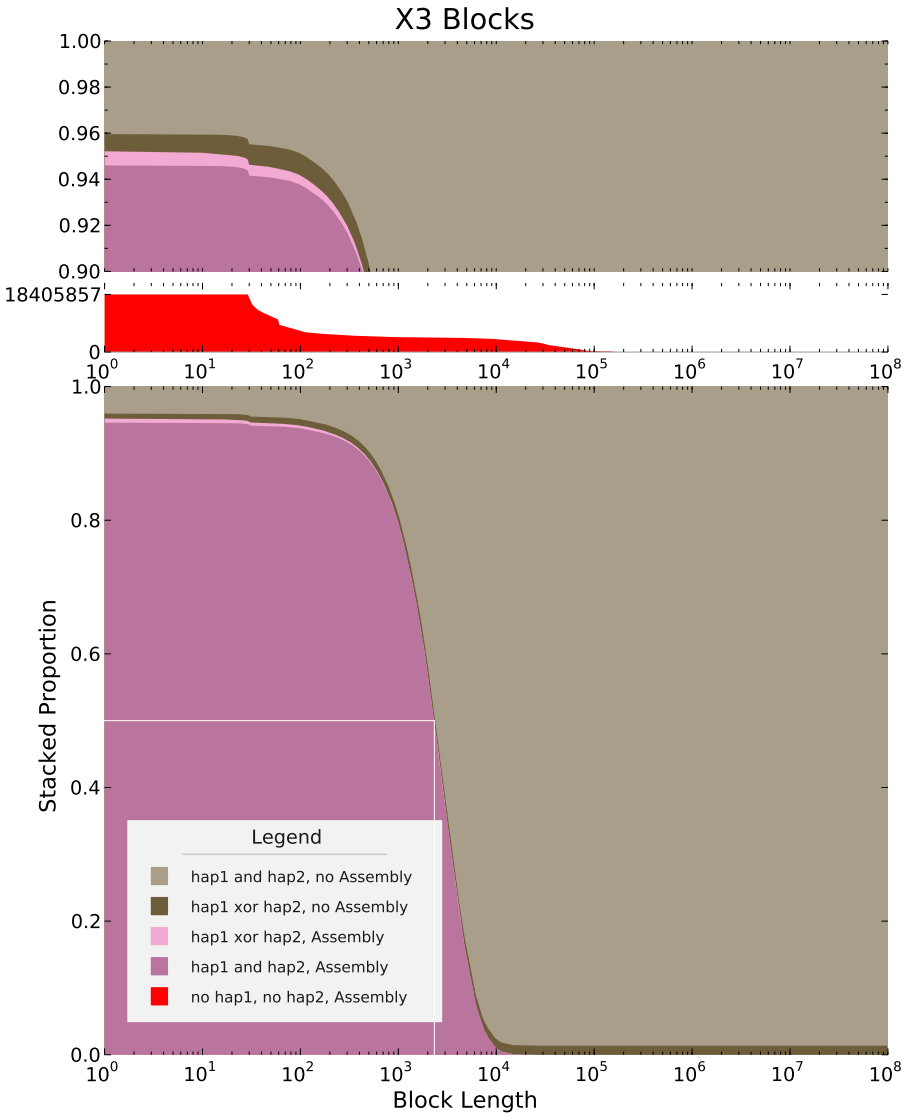


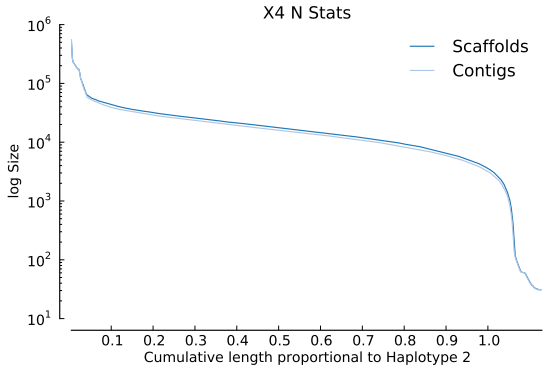
Figure 3.242: X3 blocks caption goes here.

X4

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
X5	0.96766	0.96789	0.96742	0.99789
X4	0.96758	0.96779	0.96735	0.99798
W3	0.96751	0.96771	0.96728	0.99810

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	173,372	31	32.00	39	732.71	61.00	541,257	4,152.74	127,031,925
Contigs	175,163	31	32.00	39	724.68	61.00	541,257	3,990.20	126,937,664

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	104,969,117 – 106,005,587	102,366,017 – 103,037,469	204,725,177.5 – 206,065,327.9	1,243 – 1,561
Heterozygous	408,857 – 421,961	395,949 – 402,310	781,571.0 – 793,751.8	5,071 – 5,283
Indel	1,766,601 – 2,132,972	718,895 – 954,346	1,435,522.4 – 1,904,576.8	1,070 – 1,155

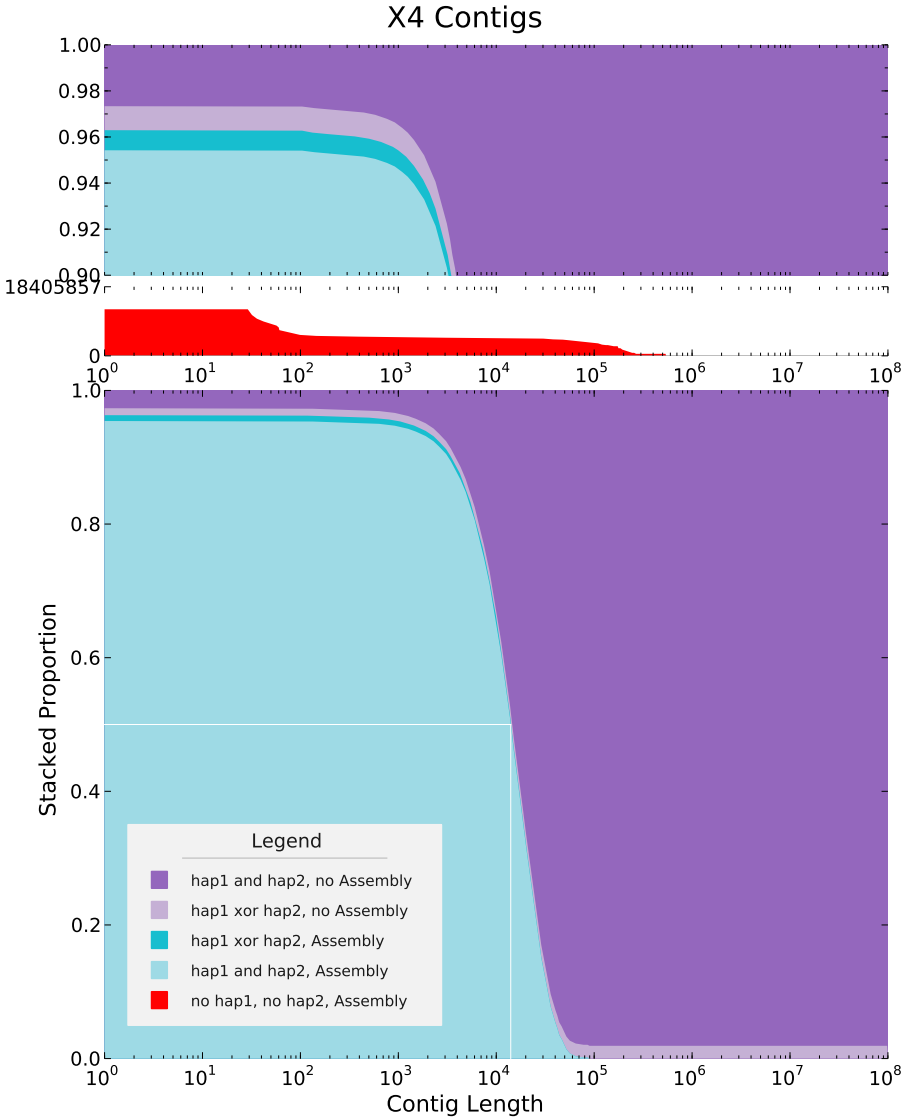


Figure 3.243: X4 contigs caption goes here.

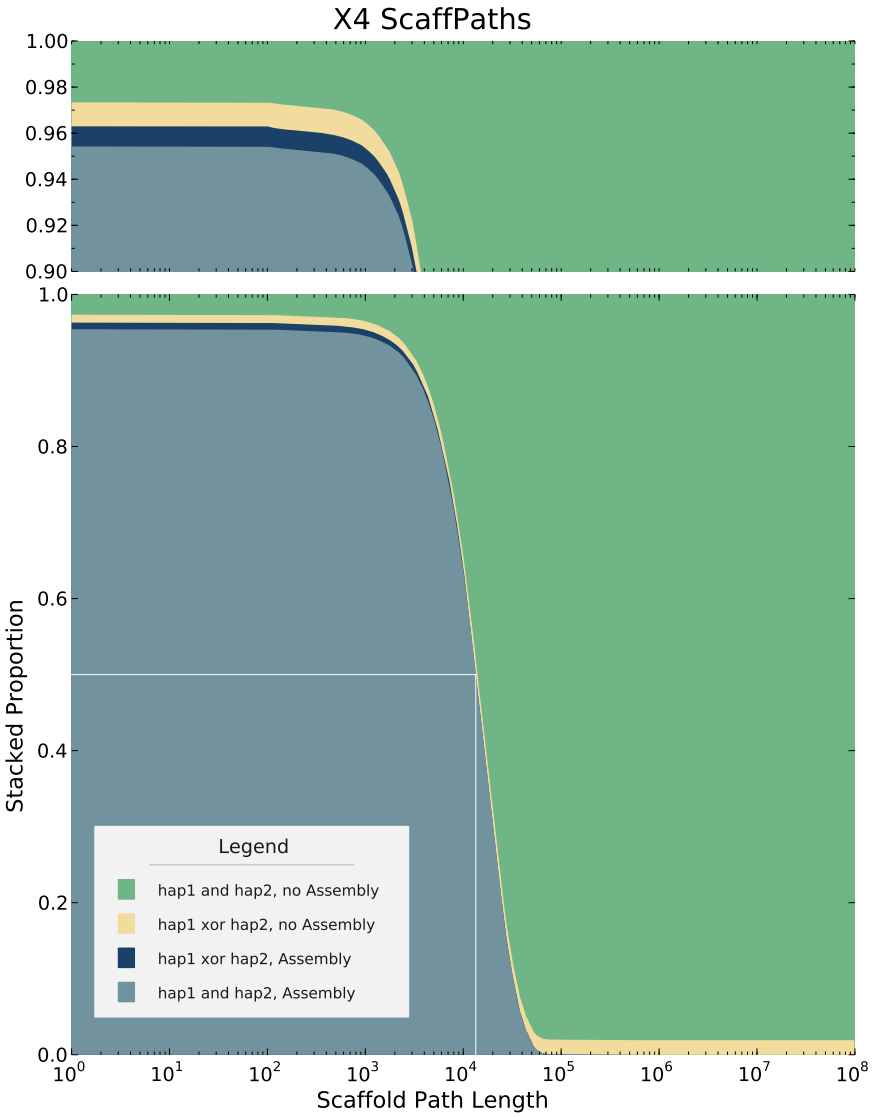


Figure 3.244: X4 scaffolds caption goes here.

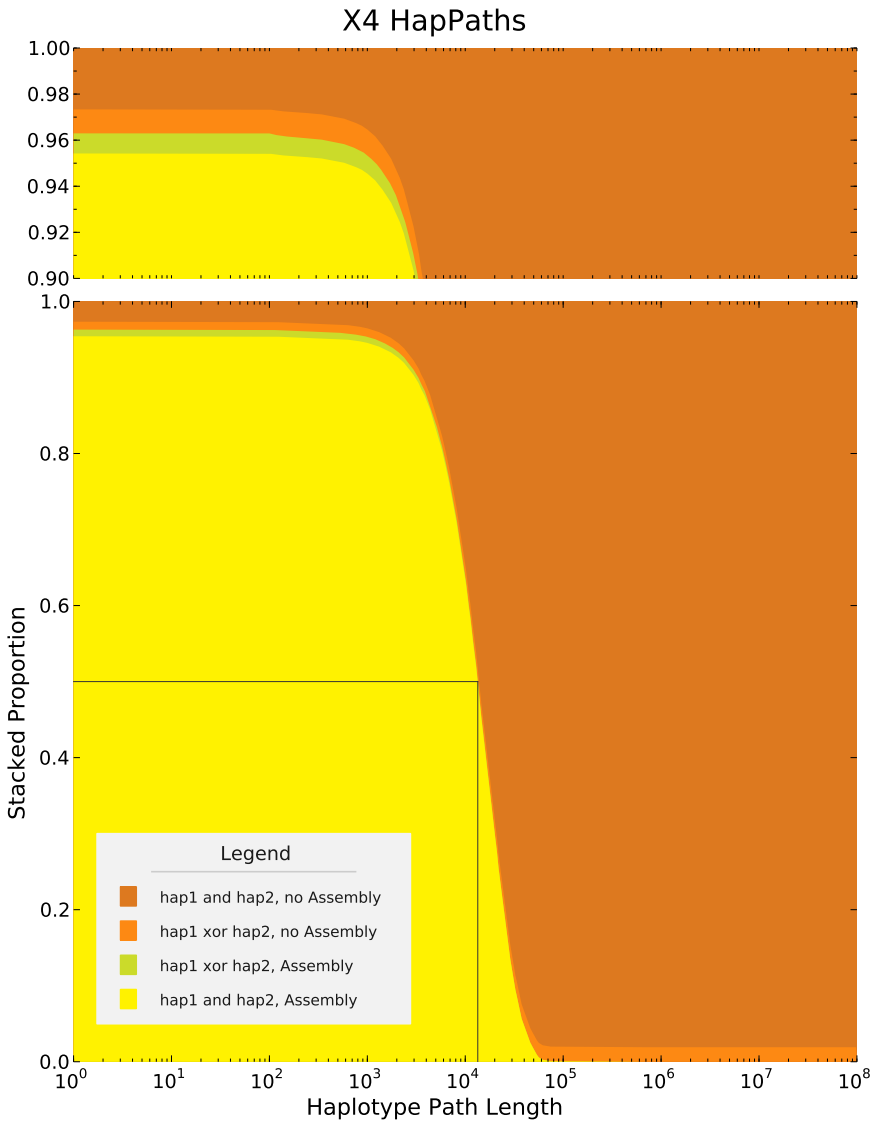


Figure 3.245: X4 hapPaths caption goes here.

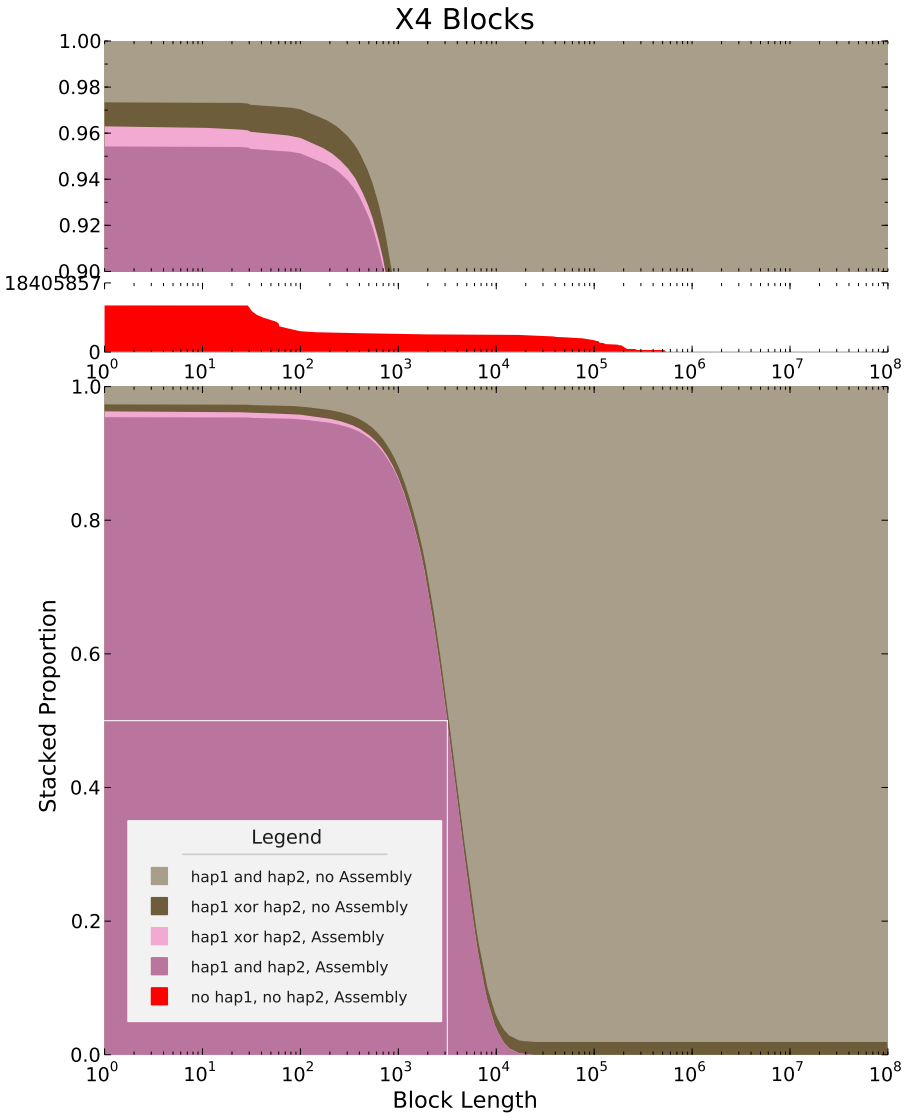


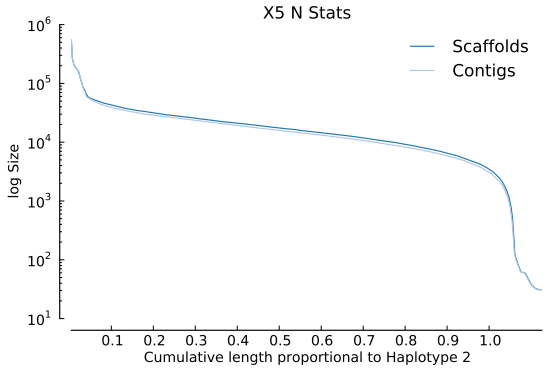
Figure 3.246: X4 blocks caption goes here.

X5

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
W10	0.96812	0.96852	0.96772	0.99812
X5	0.96766	0.96789	0.96742	0.99789
X4	0.96758	0.96779	0.96735	0.99798

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	172,923	31	32.00	39	732.53	61.00	541,257	4,119.62	126,671,540
Contigs	174,679	31	32.00	39	724.65	61.00	541,257	3,964.12	126,580,753

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	105,280,830 – 106,319,236	102,682,740 – 103,354,269	205,358,873.9 – 206,699,207.7	1,185 – 1,478
Heterozygous	409,874 – 423,099	396,971 – 403,405	783,980.8 – 796,343.2	4,894 – 5,087
Indel	1,768,904 – 2,134,099	729,569 – 964,024	1,456,897.4 – 1,924,004.8	1,053 – 1,144

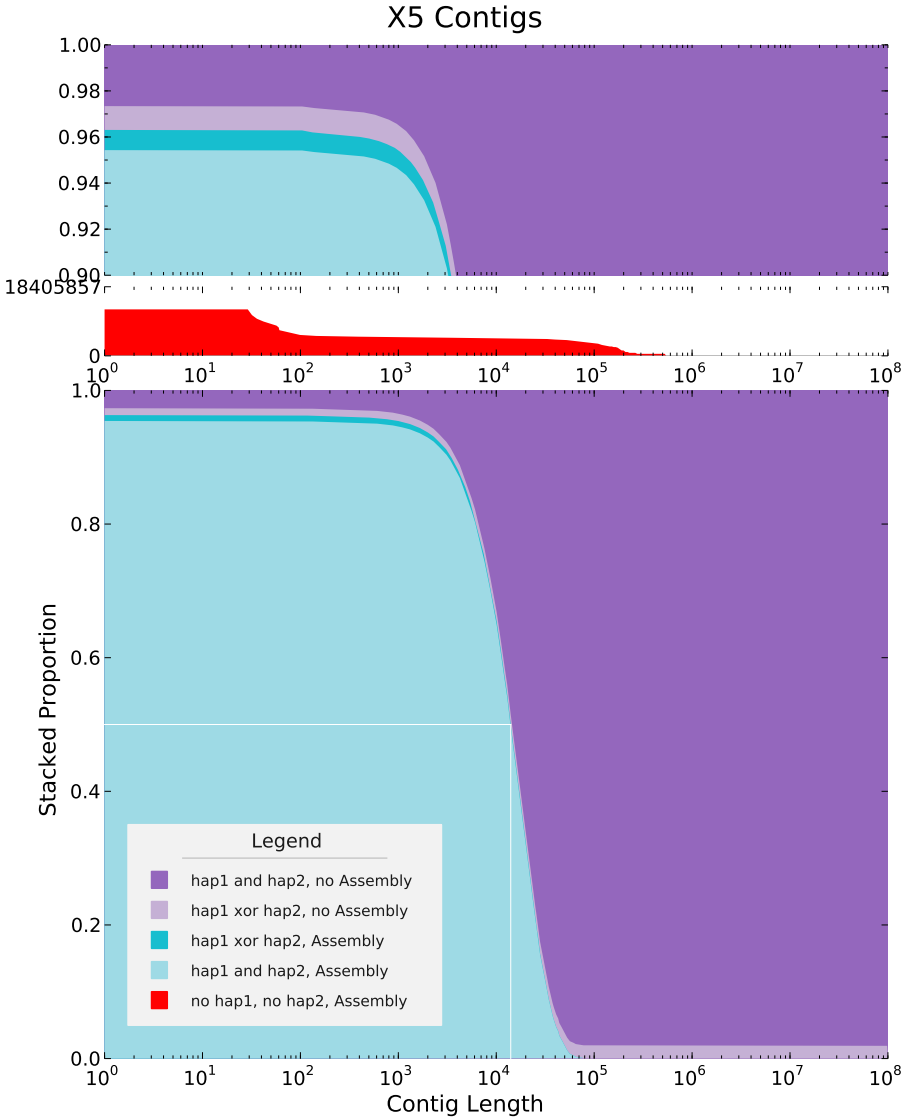


Figure 3.247: X5 contigs caption goes here.

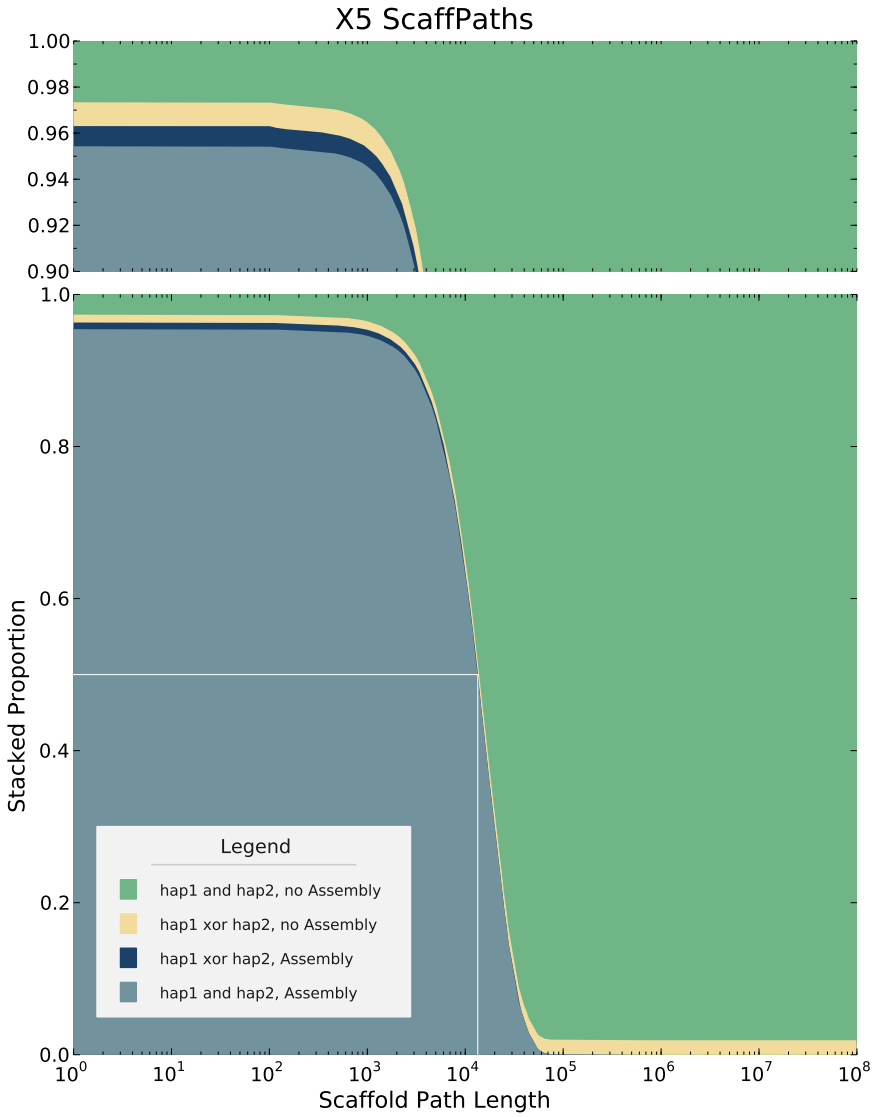


Figure 3.248: X5 scaffolds caption goes here.

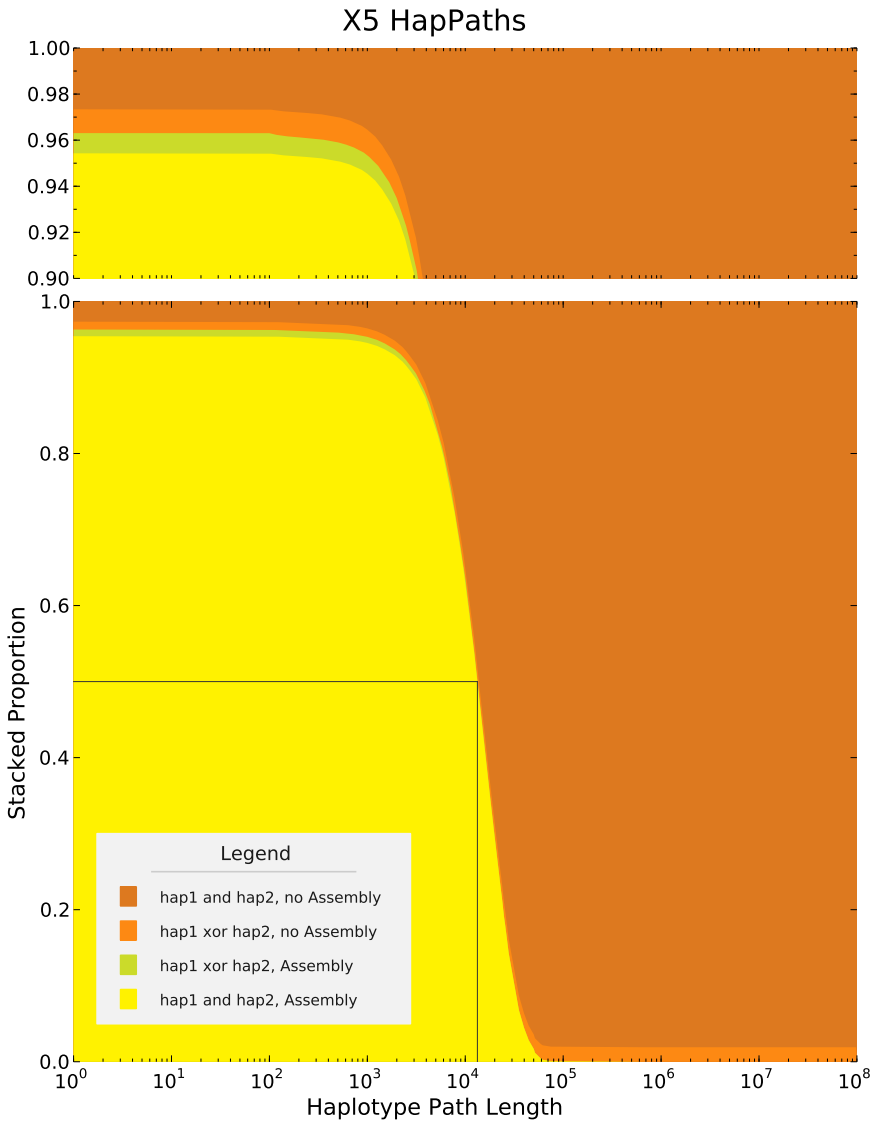


Figure 3.249: X5 hapPaths caption goes here.

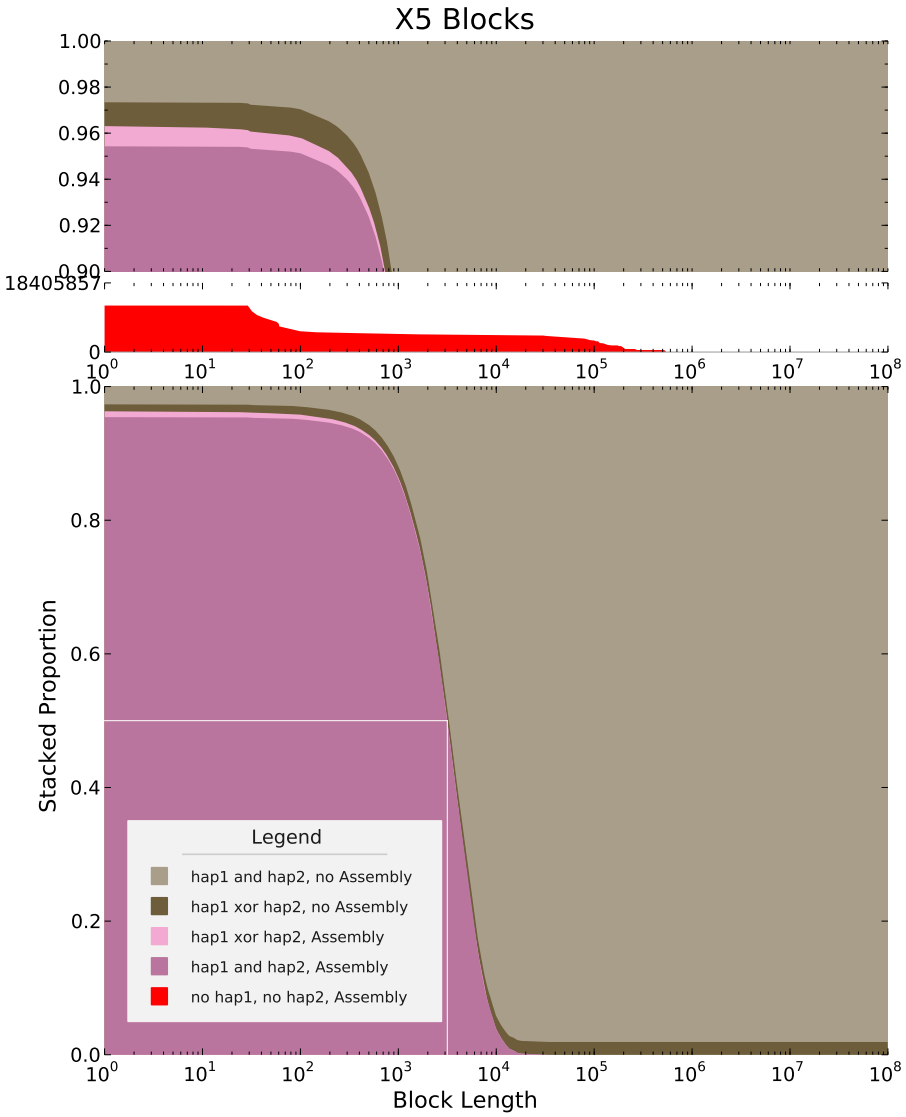


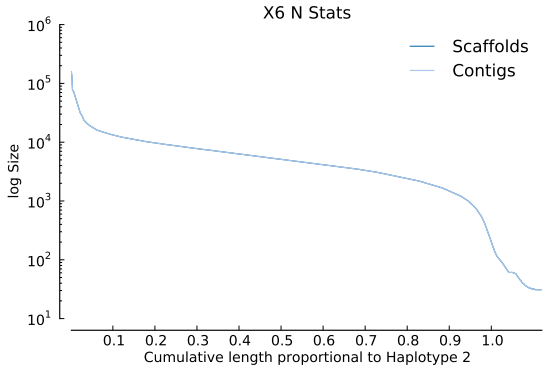
Figure 3.250: X5 blocks caption goes here.

X6

Coverage neighbor table:

ID	Total	Hap 1	Hap 2	Bac
E3	0.95559	0.95601	0.95517	0.99317
X6	0.95516	0.95527	0.95504	0.99562
X3	0.95516	0.95535	0.95495	0.99560

Submitted assembly N stats plot



Submitted assembly size stats table

Category	n	min	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev	Sum
Scaffolds	311,185	31	32.00	41	405.06	68.00	156,082	1,600.42	126,049,808
Contigs	311,185	31	32.00	41	405.06	68.00	156,082	1,600.42	126,049,808

SNP stats table

Category	Total	Calls	Correct (bits)	Errors
Homozygous	108,437,056 – 110,212,352	104,681,739 – 105,744,241	209,363,452.0 – 211,488,378.0	13 – 52
Heterozygous	412,330 – 436,617	394,658 – 401,317	789,316.0 – 802,634.0	0 – 0
Indel	1,185,234 – 1,558,748	455,655 – 691,637	910,146.0 – 1,381,984.0	582 – 645

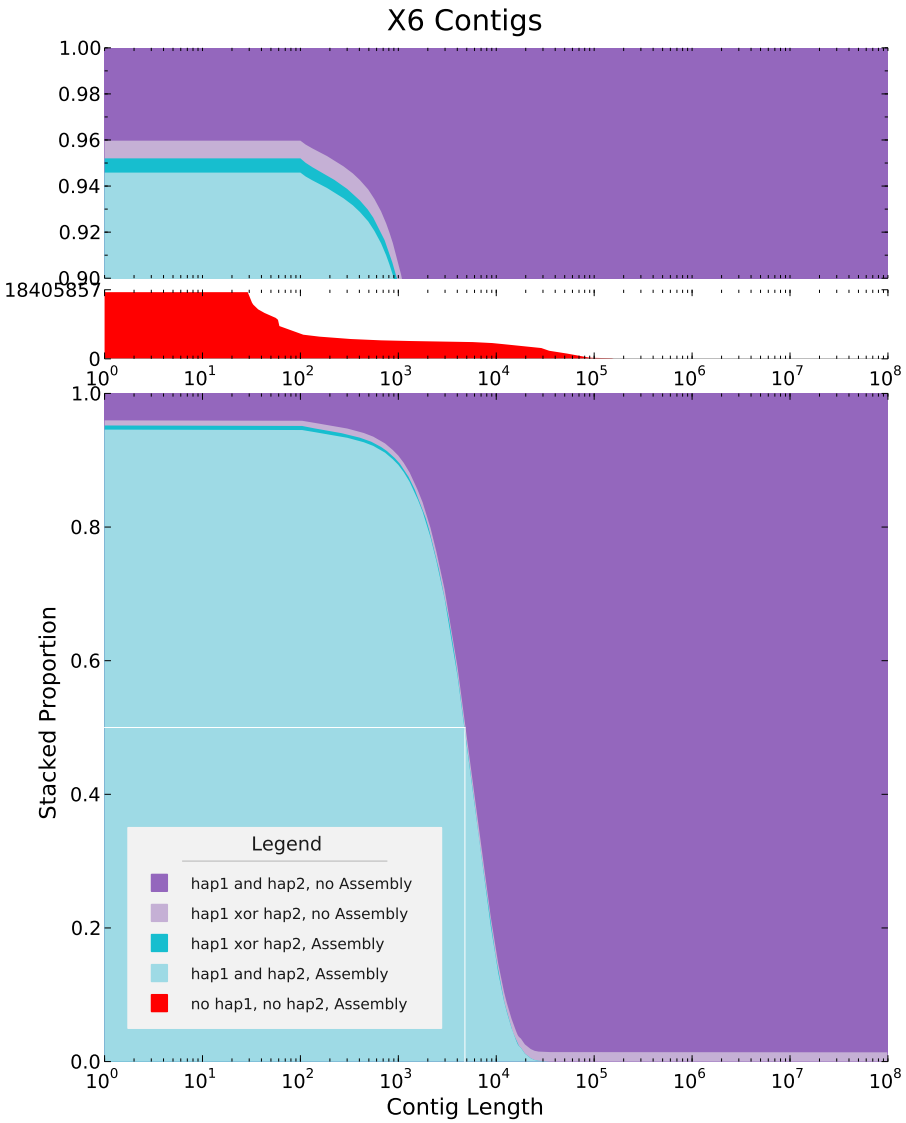


Figure 3.251: X6 contigs caption goes here.

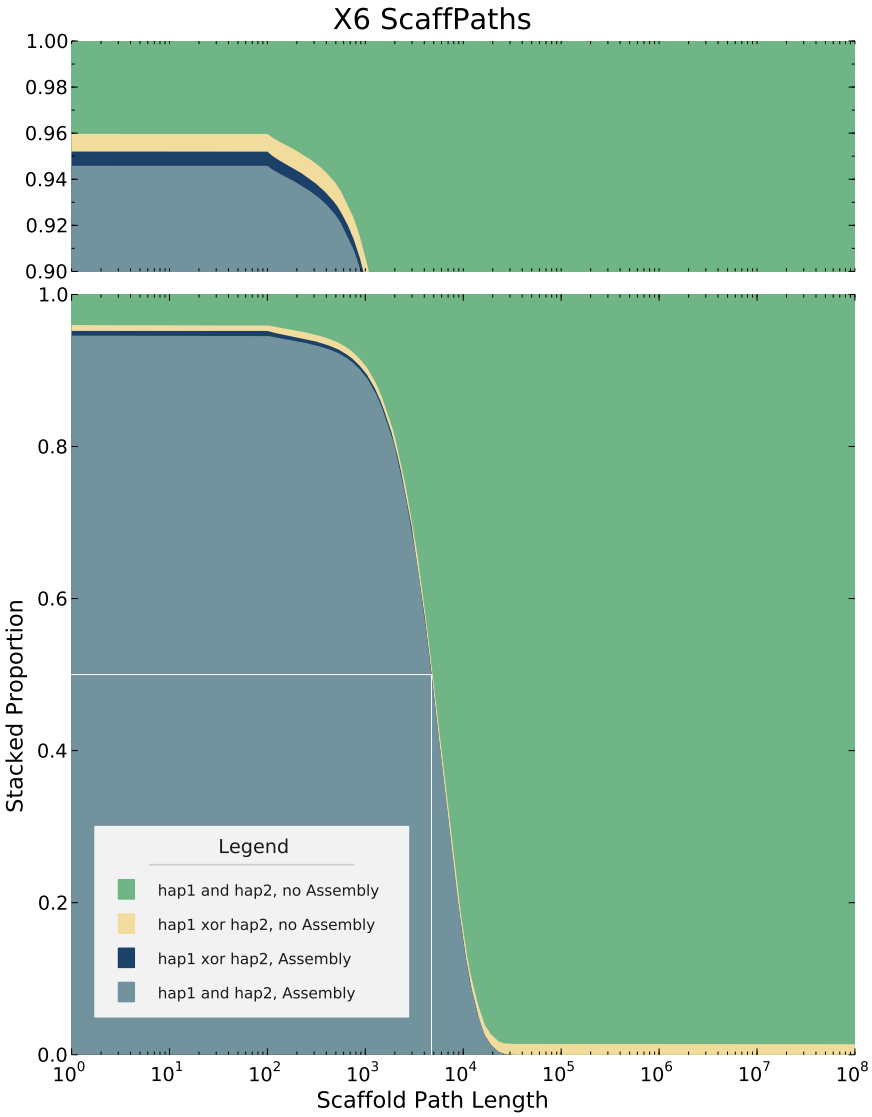


Figure 3.252: X6 scaffolds caption goes here.

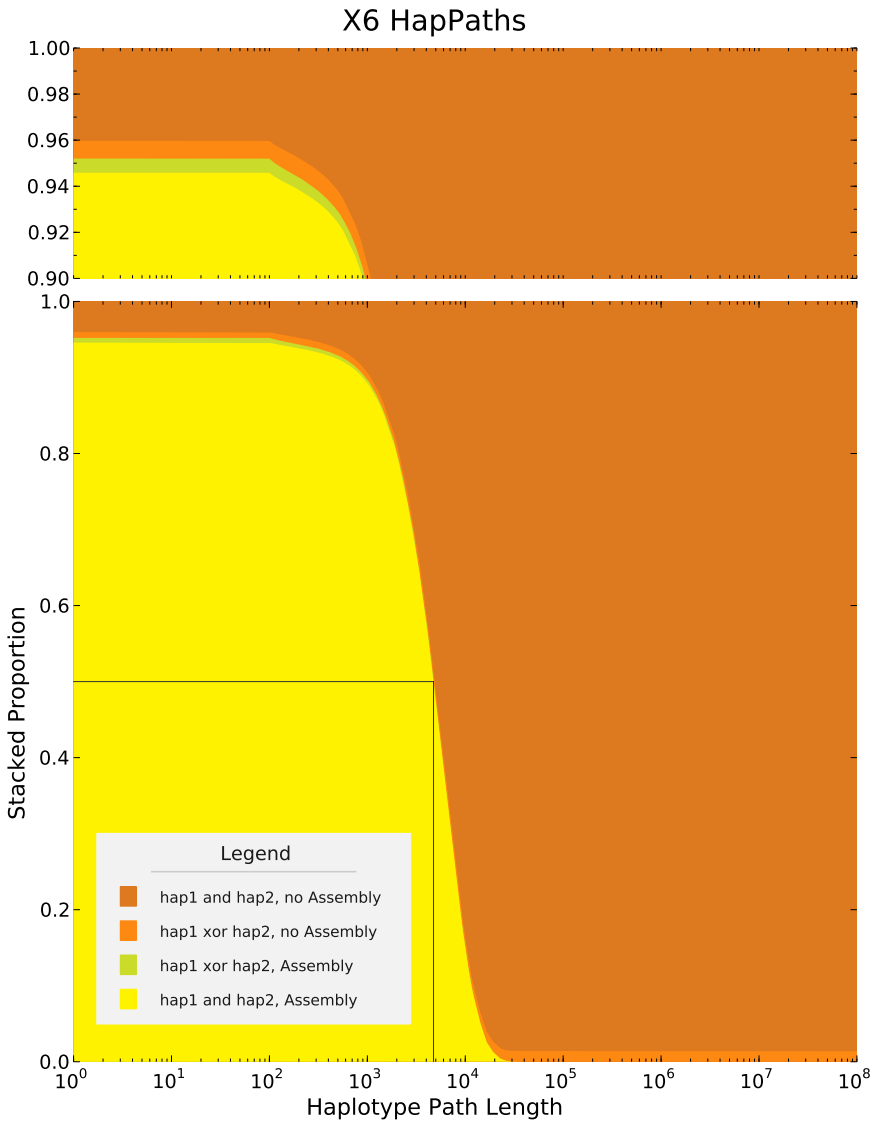


Figure 3.253: X6 hapPaths caption goes here.

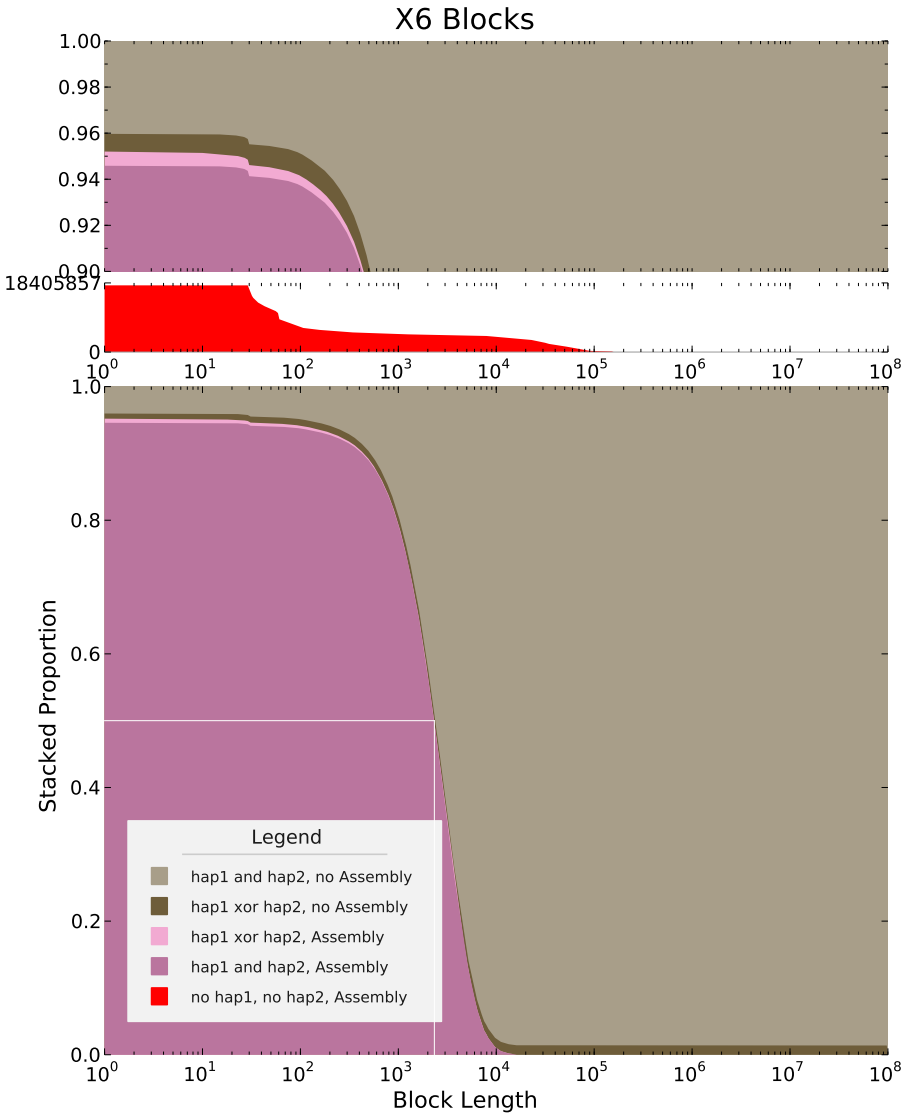


Figure 3.254: X6 blocks caption goes here.

Bibliography

[Alekseyev and Pevzner, 2007] Alekseyev, M. and Pevzner, P., 2007. Whole genome duplications and contracted breakpoint graphs. *SIAM JOURNAL ON COMPUTING*, .

[Benson et al., 1999] Benson, D., Boguski, M., Lipman, D., Ostell, J., Ouellette, B. F., Rapp, B., and Wheeler, D. A., 1999. GenBank. *Nucleic Acids Research*, **27**(1):12–17.

[Blanchette et al., 2004] Blanchette, M., Kent, W. J., Riemer, C., Elnitski, L., Smit, A. F. A., Roskin, K. M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E. D., *et al.*, 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*, **14**(4):708–15.

[Edgar et al., 2010] Edgar, R. C., Asimenos, G., Batzoglou, S., and Sidow, A., Last accessed July 2010. Evolver: a whole-genome sequence evolution simulator. *http://www.drive5.com/evolver/*, .

[Fujita et al., 2010] Fujita, P. A., Rhead, B., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Cline, M. S., Goldman, M., Barber, G. P., Clawson, H., Coelho, A., *et al.*, 2010. The UCSC Genome Browser database: update 2011. *Nucleic Acids Research*, **39**(Database):D876–D882.

[Medvedev and Brudno, 2009] Medvedev, P. and Brudno, M., 2009. Maximum likelihood genome assembly. *J Comput Biol*, **16**(8):1101–16.

[Paten et al., 2011] Paten, B., Diekhans, M., Earl, D., John, J. S., Ma, J., Suh, B., and Haussler, D., 2011. Cactus graphs for genome comparisons. *J Comput Biol*, **18**(3):469–81.

[Paten et al.,] Paten, B., Earl, D., Nguyen, N., Diekhans, M., Zerbino, D., and Haussler, D. Cactus: algorithms for genome multiple sequence alignment. *submitted*, .

[Smit, 1999] Smit, A. F., 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Current opinion in genetics & development*, **9**(6):657–663.

[Zerbino and Birney, 2008] Zerbino, D. R. and Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de bruijn graphs. *Genome Res*, **18**(5):821–9.