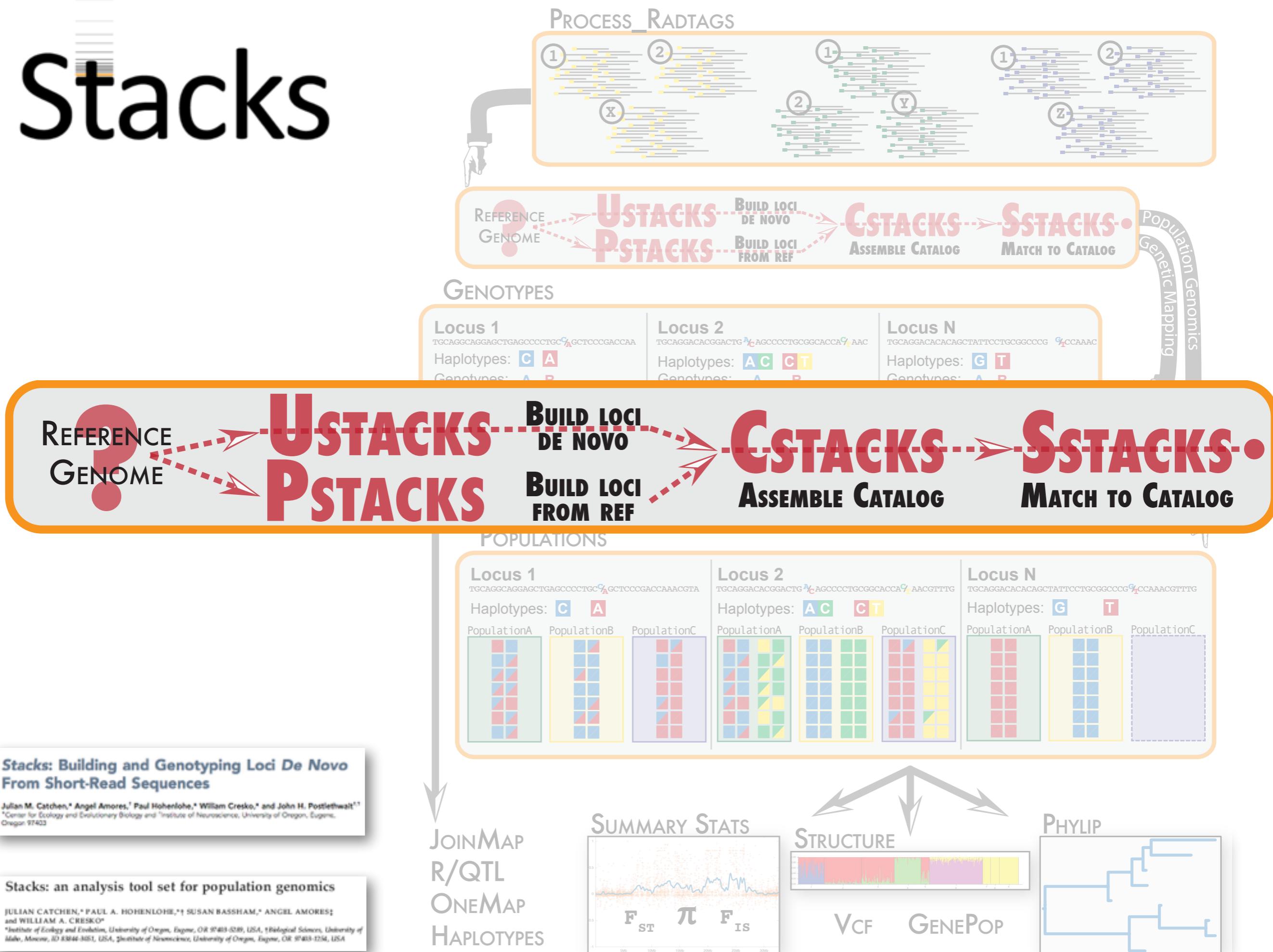
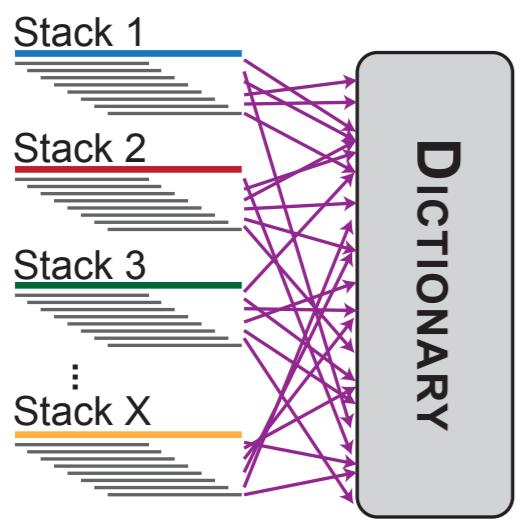
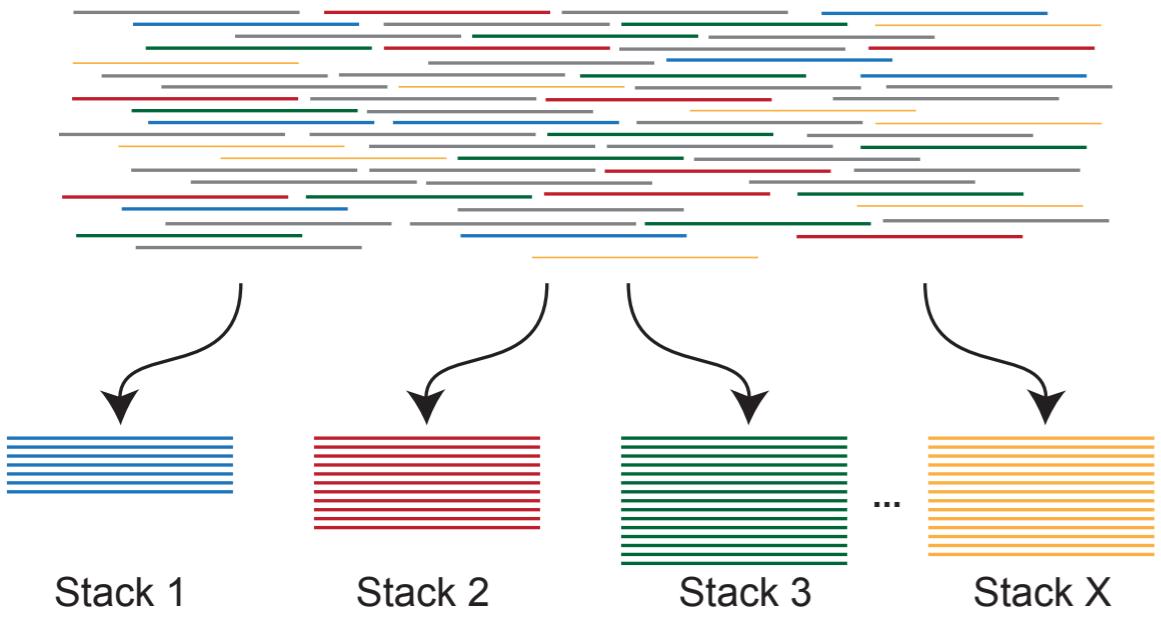


Stacks





Matching stacks using k-mers

A string can

- I. Replace an n^2 algorithm with linear hash table lookup
 2. Further reduce the search space:
 - I. Given a number of allowed mismatches calculate the maximum useful k-mer size
 2. For each sequence: lookup matching k-mers and record associated reads
 3. For pairs of sequences with minimum number of k-mers, align reads.

TAC**CC**ACATC
TAC**CC**ACATC
ACCACATC
CCACATC
CACATC
ACATC
CATC
ATC
TC
C

AATCCTT

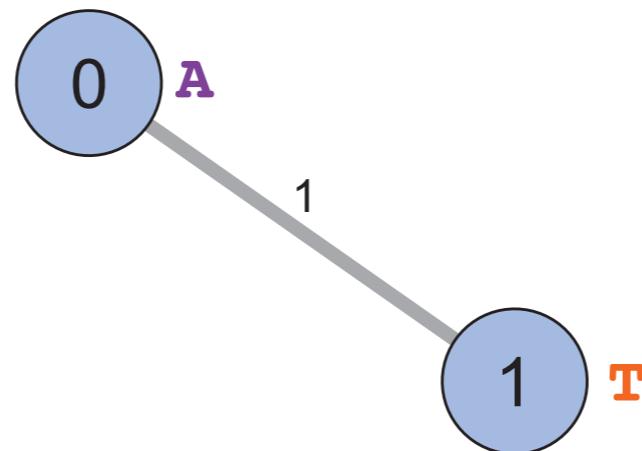
um
A
AA
AAA
AAAT
AAATC
ACGTGTGAAAAATCC
CGTGTGAAAAATCCT
GTGTGAAAAATCCTT
TGTGAAAAATCCTTT

ACGTGTGAAAAATCC
CGTGTGAAAAATCCT
GTGTGAAAAATCCTT
TGTGAAAAATCCTTT

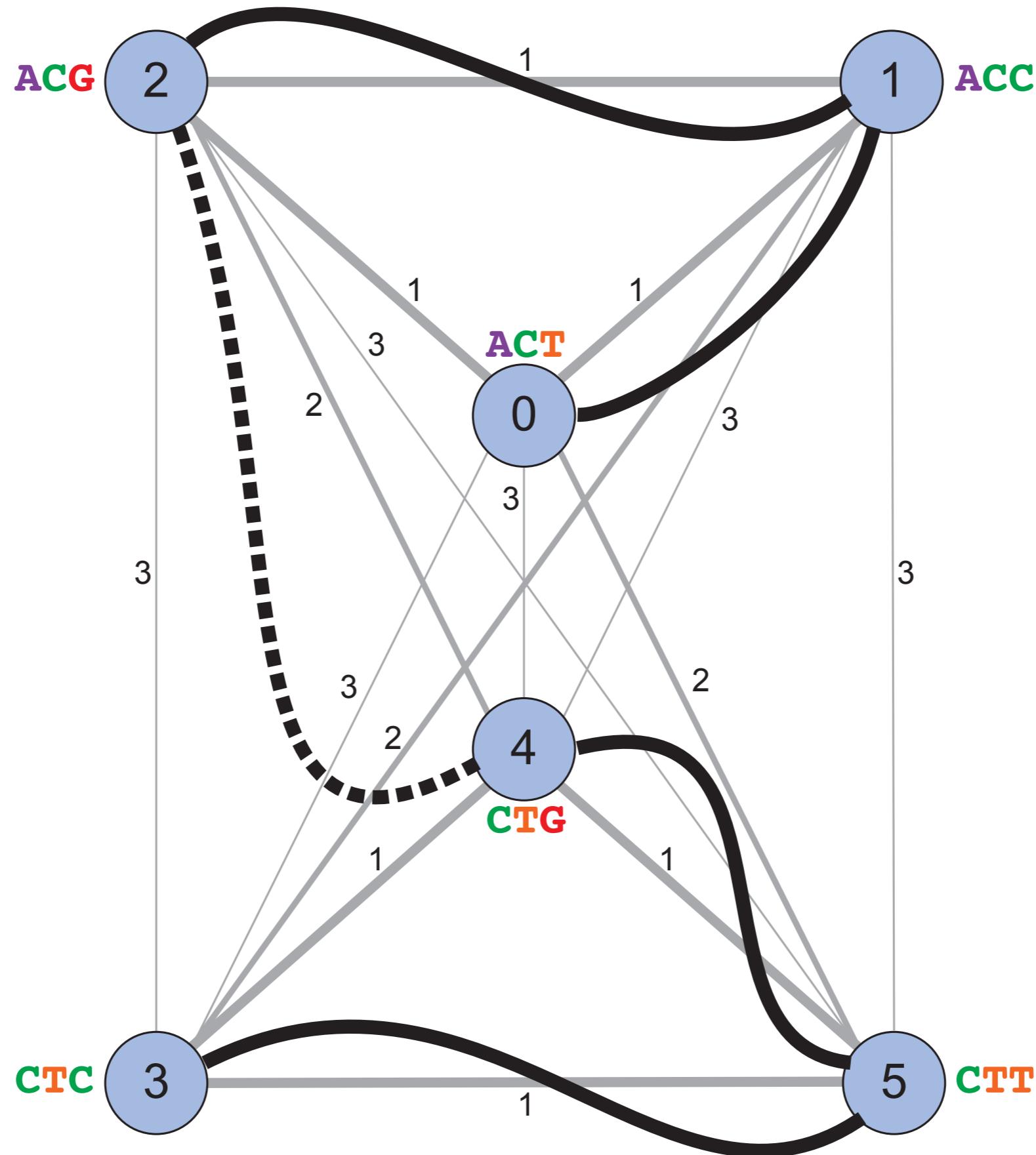
Building loci using minimum spanning trees

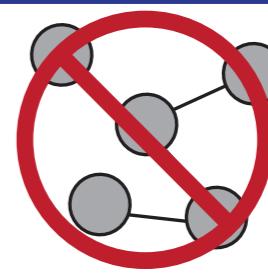
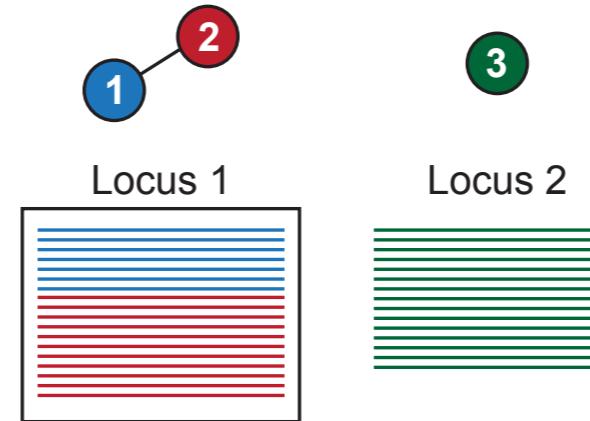
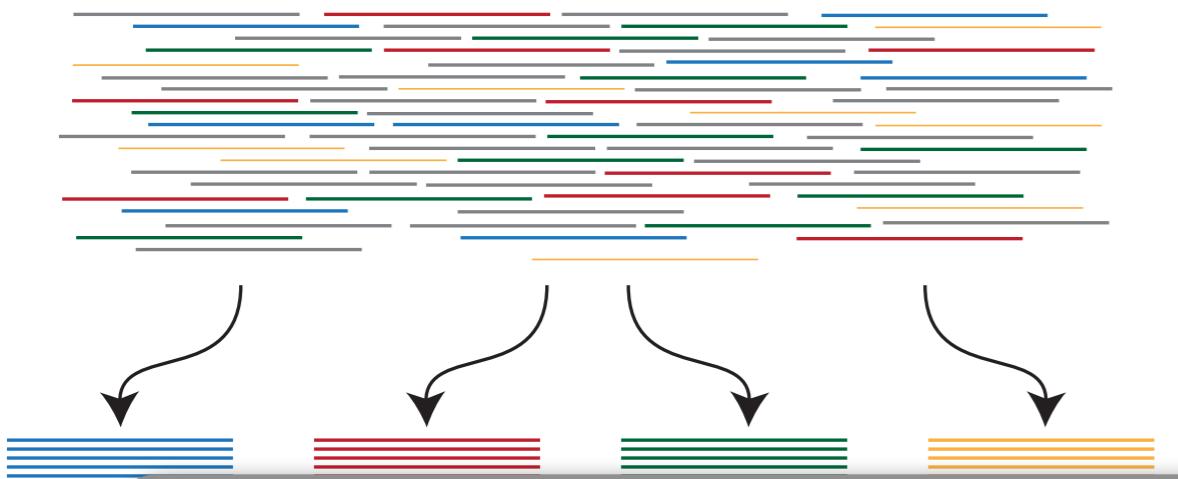
TGCAGGTACCCCGTCCCC**A**GC GGAGGACCTGTTACC
TGCAGGTACCCCGTCCCC**A**GC GGAGGACCTGTTACC
TGCAGGTACCCCGTCCCC**A**GC GGAGGACCTGTTACC
TGCAGGTACCCCGTCCCC**A**GC GGAGGACCTGTTACC
TGCAGGTACCCCGTCCCC**A**GC GGAGGACCTGTTACC

TGCAGGTACCCCGTCCCC**T**GC GGAGGACCTGTTACC
TGCAGGTACCCCGTCCCC**T**GC GGAGGACCTGTTACC
TGCAGGTACCCCGTCCCC**T**GC GGAGGACCTGTTACC
TGCAGGTACCCCGTCCCC**T**GC GGAGGACCTGTTACC
TGCAGGTACCCCGTCCCC**T**GC GGAGGACCTGTTACC



Building loci using minimum spanning trees





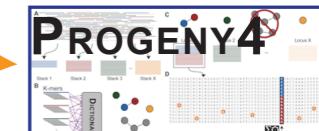
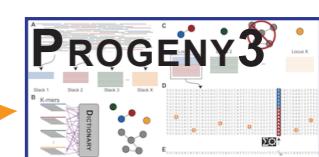
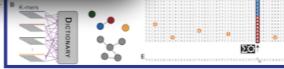
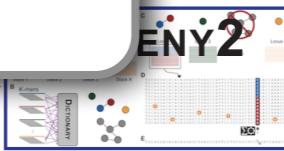
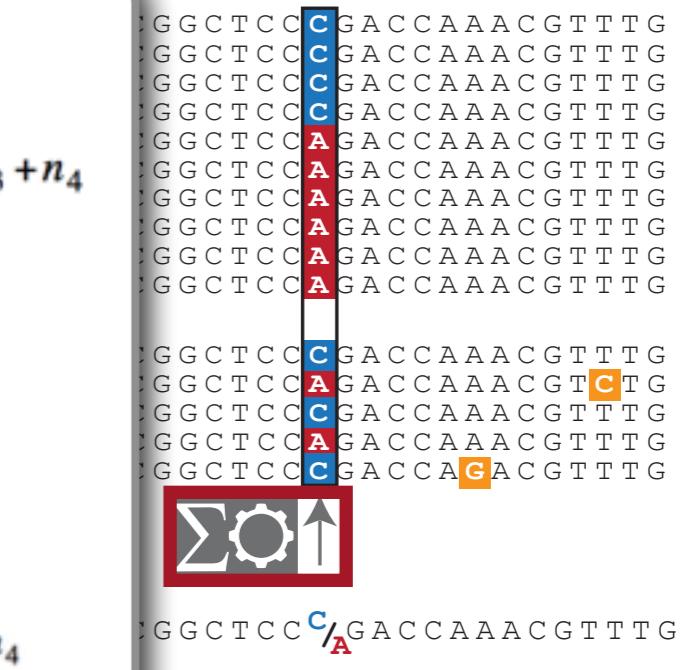
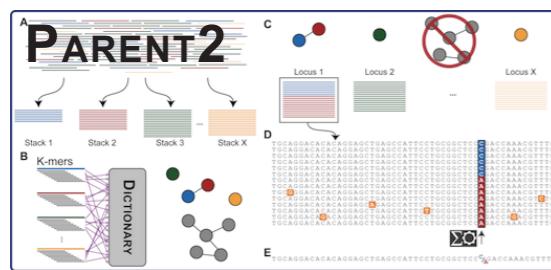
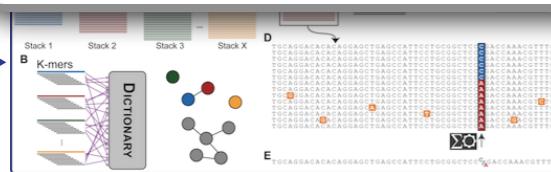
Locus X

Sta Homozygote Likelihood

$$L(1/1) = P(n_1, n_2, n_3, n_4 | 1/1) = \frac{n!}{n_1! n_2! n_3! n_4!} \left(1 - \frac{3\varepsilon}{4}\right)^{n_1} \left(\frac{\varepsilon}{4}\right)^{n_2 + n_3 + n_4}$$

Heterozygote Likelihood

$$L(1/2) = P(n_1, n_2, n_3, n_4 | 1/2) = \frac{n!}{n_1! n_2! n_3! n_4!} \left(0.5 - \frac{\varepsilon}{4}\right)^{n_1+n_2} \left(\frac{\varepsilon}{4}\right)^{n_3+n_4}$$



Differentiating SNPs from error

Restriction enzyme recognition site



TATAGCTGCATTTCATGTTCATTTATATTGAGTGAAGTTTCAGCCCTGCAGGTCTCCTGCTACCACCAATTCCACCTCTGCAGCTCATCCACAATGCAGCAGACCG
CTGCATTTCATGTTCATTTATATTGAGTGAAGTTTCAGCCCTGCA
CTGCATTTCATGTTCATTTATATTGAGTGAAGTATTCAAGCCCTGCA
CTGCATTTCATGTTCATTTATATTGAGTGAAGTTTCAGCCCTGCA
CTGCATTGCAAGTTCATTTATATTGAGTGAAGTTTCAGCCCTGCA
CTGCATTTCATGTTCATTTATATTGAGTGAAGTTTCAGCCCTGCA
CTGCATTTCATGTTCATTTATATTGAGTGAAGTATTCAAGCCCTGCA
CTGCATTTCATGTTCACTTCATTTATATTGAGTGAAGTTTCAGCCCTGCA
CTGCATTTCATGTTCATTTATATTGAGTGAAGTTTCAGCCCTGCA
CTGCATTTCATGTTCATTTATATTGAGTGAAGTTTCAGCCCTGCA

Reference Genome



Sequence Reads

Differentiating SNPs from error

TATAGCTGCATTCATGTT
CTGCATTCATGTT
CTGCATTCATGTT
CTGCATTCATGTT
CTGCATTGATGTT
CTGCATTCATGTT
CTGCATTCATGTT
CTGCATTCATGTT
CTGCATTCATGTT
CTGCATTCATGTT

$L(n_1 \text{ hom}) = P(n_1, n_2, n_3, n_4) = \frac{n!}{n_1! n_2! n_3! n_4!} \left(1 - \frac{3\epsilon}{4}\right)^{n_1} \left(\frac{\epsilon}{4}\right)^{n_2} \left(\frac{\epsilon}{4}\right)^{n_3} \left(\frac{\epsilon}{4}\right)^{n_4}$

$L(n_1 n_2 \text{ het}) = P(n_1, n_2, n_3, n_4) = \frac{n!}{n_1! n_2! n_3! n_4!} \left(0.5 - \frac{\epsilon}{4}\right)^{n_1} \left(0.5 - \frac{\epsilon}{4}\right)^{n_2} \left(\frac{\epsilon}{4}\right)^{n_3} \left(\frac{\epsilon}{4}\right)^{n_4}$

Maximum likelihood genotyping based on
multinomial distribution of nucleotide reads

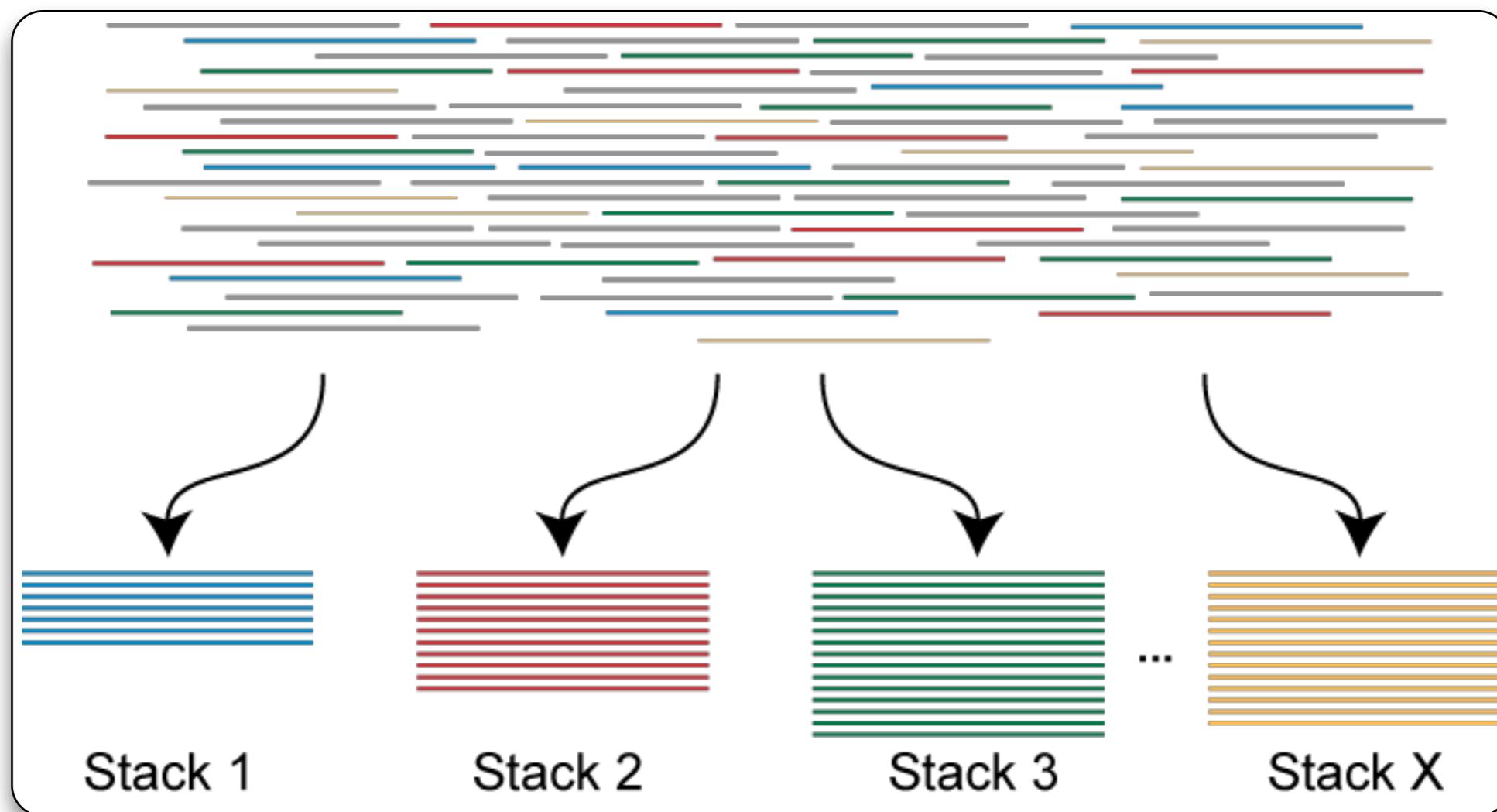
Maximum likelihood genotyping based on multinomial distribution of nucleotide reads

Parameters to control stack formation

Parameter Description	<code>denovo_map.pl</code> Parameter	Pipeline component	Component Parameter	Default Value
Minimum stack depth / minimum depth of coverage	<code>-m</code>	<code>ustacks</code>	<code>-m</code>	3
Distance allowed between stacks	<code>-M</code>	<code>ustacks</code>	<code>-M</code>	2
Distance allowed between catalog loci	<code>-n</code>	<code>cstacks</code>	<code>-n</code>	0

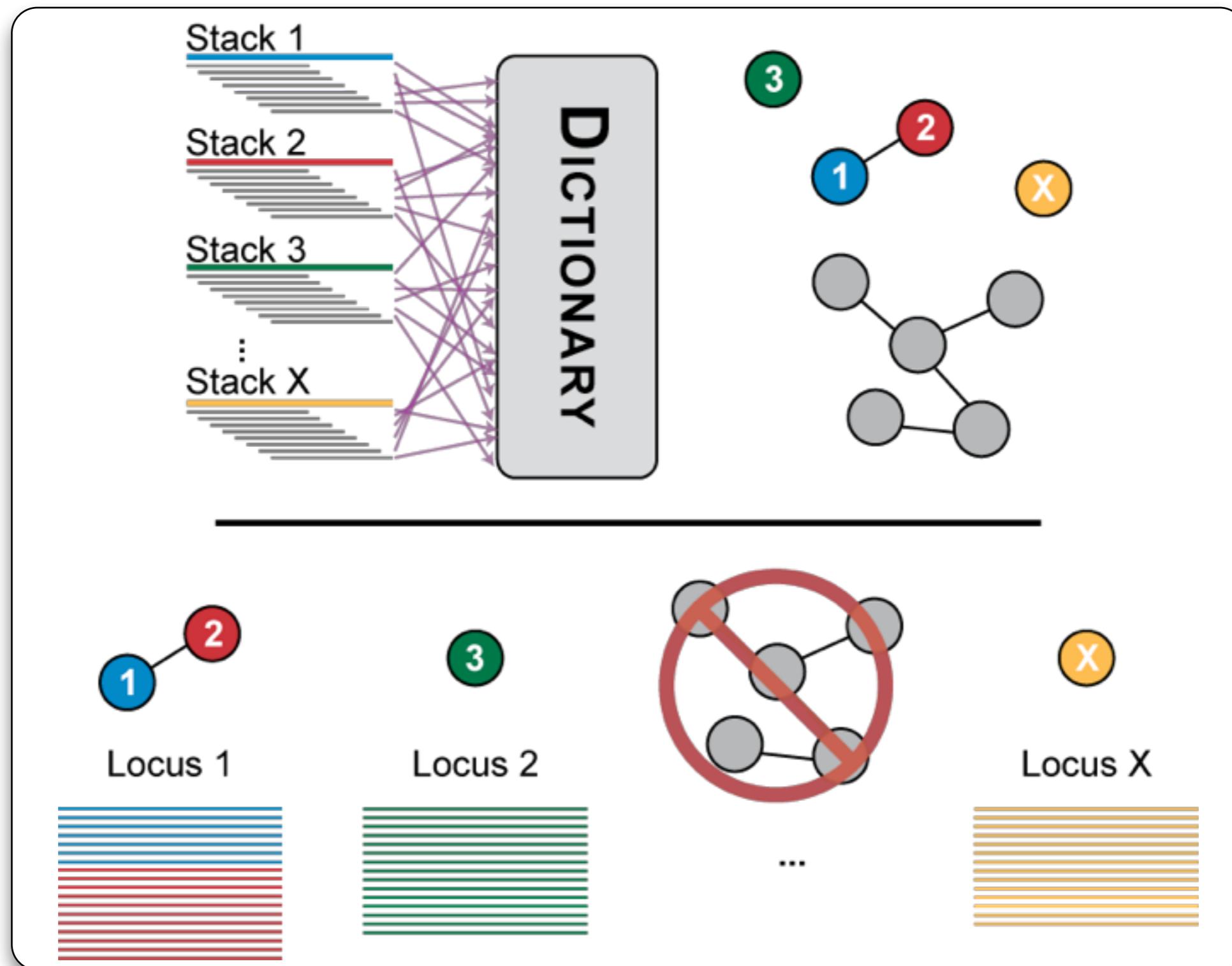
Parameters to control stack formation, ctd.

`-m <min stack depth>`



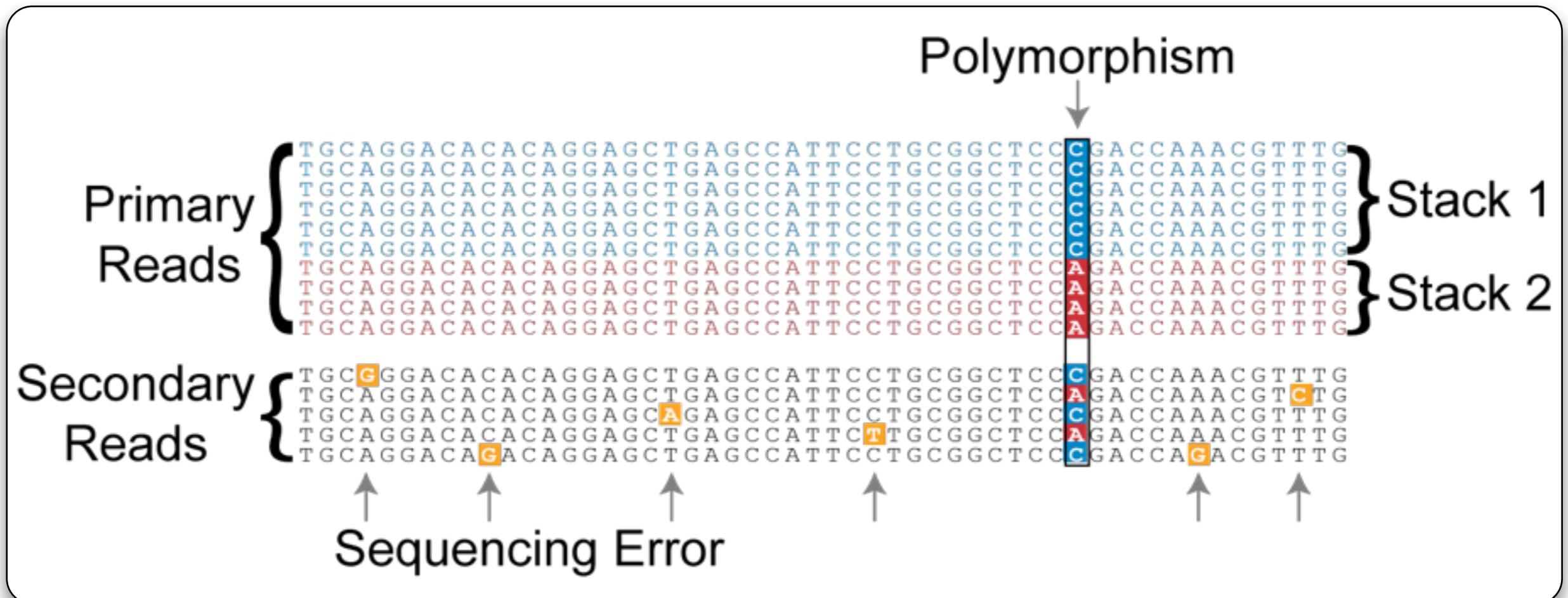
Parameters to control stack formation, ctd.

-M <distance allowed between stacks>



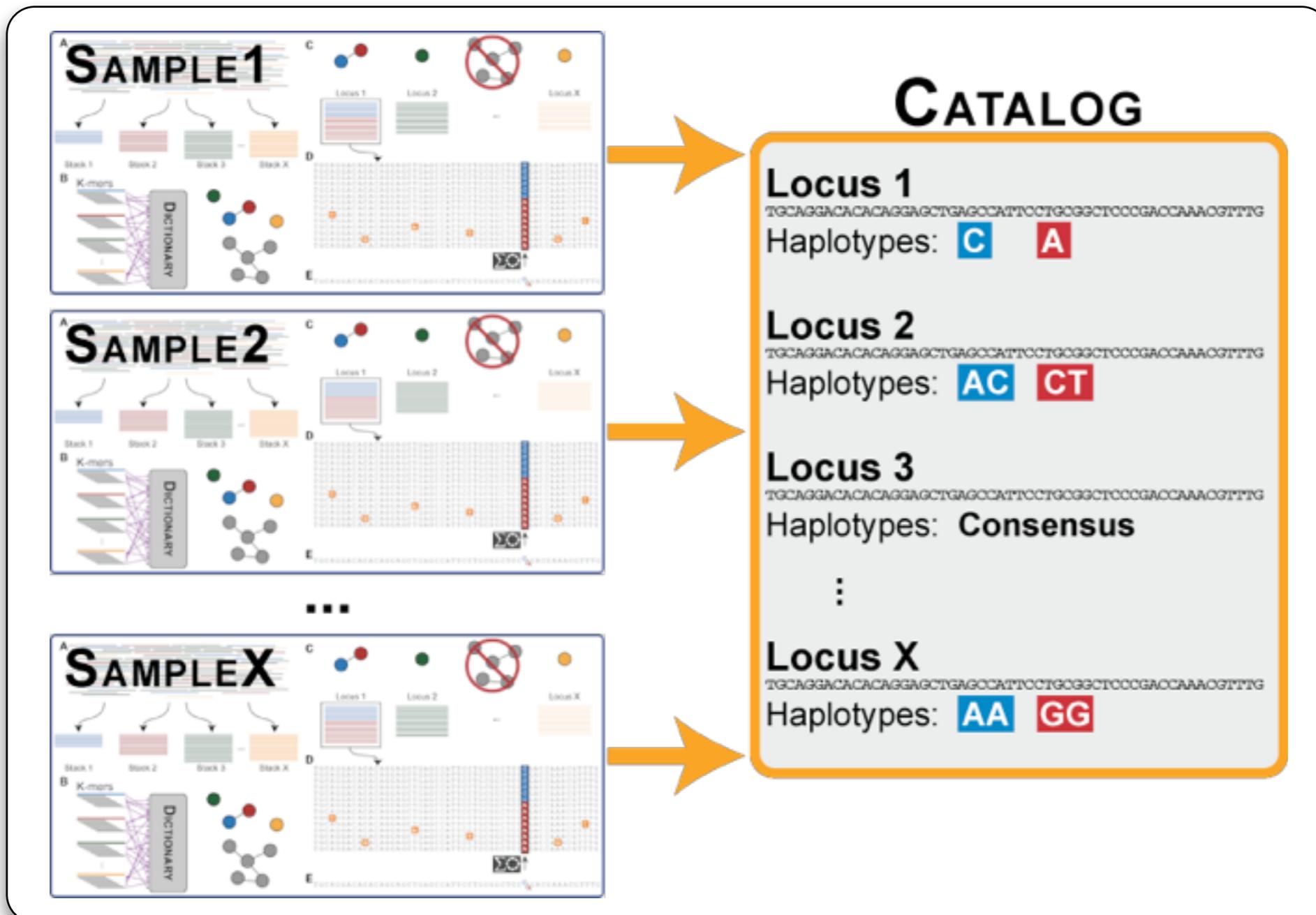
Parameters to control stack formation, ctd.

-N <secondary read distance>

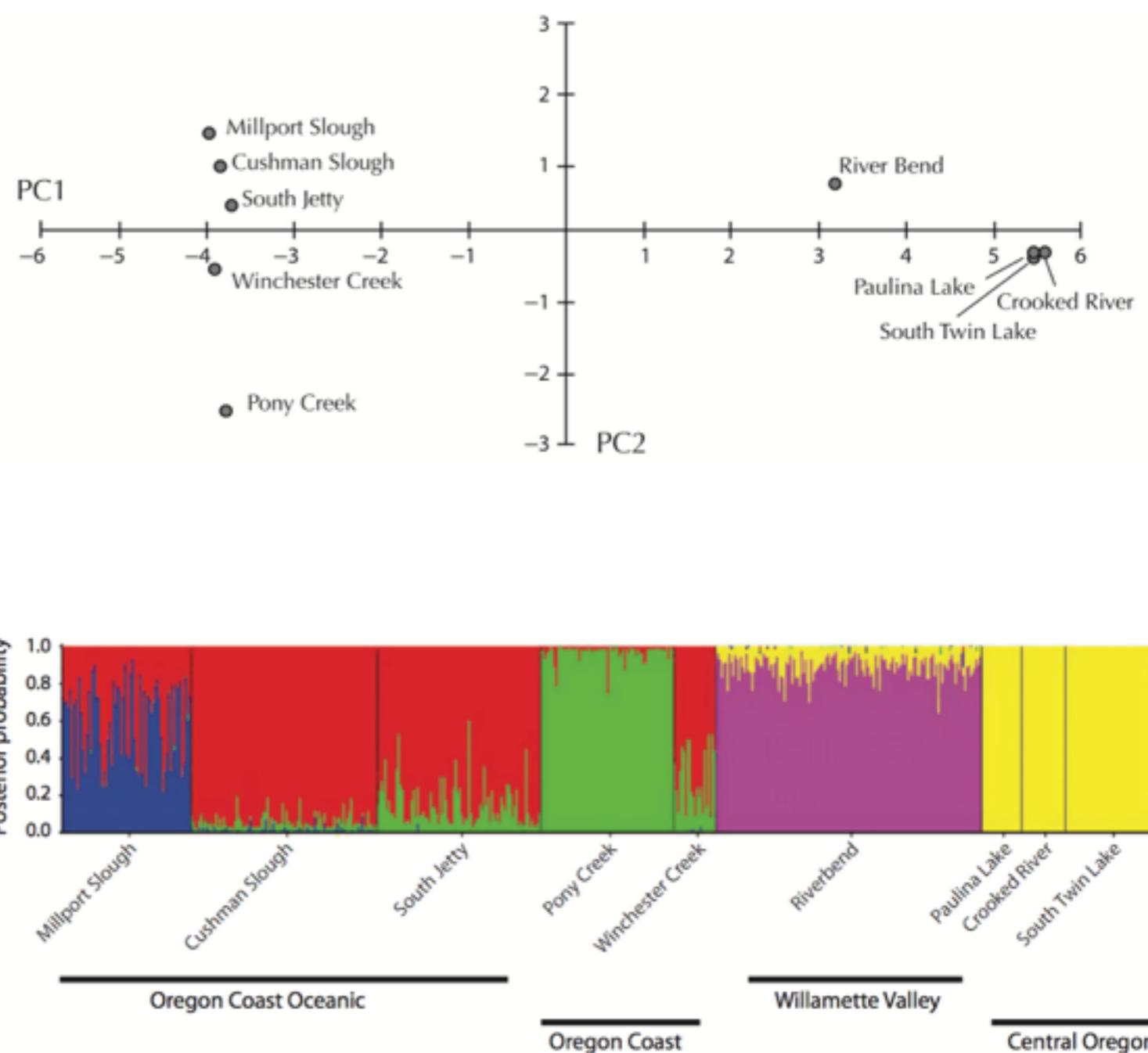
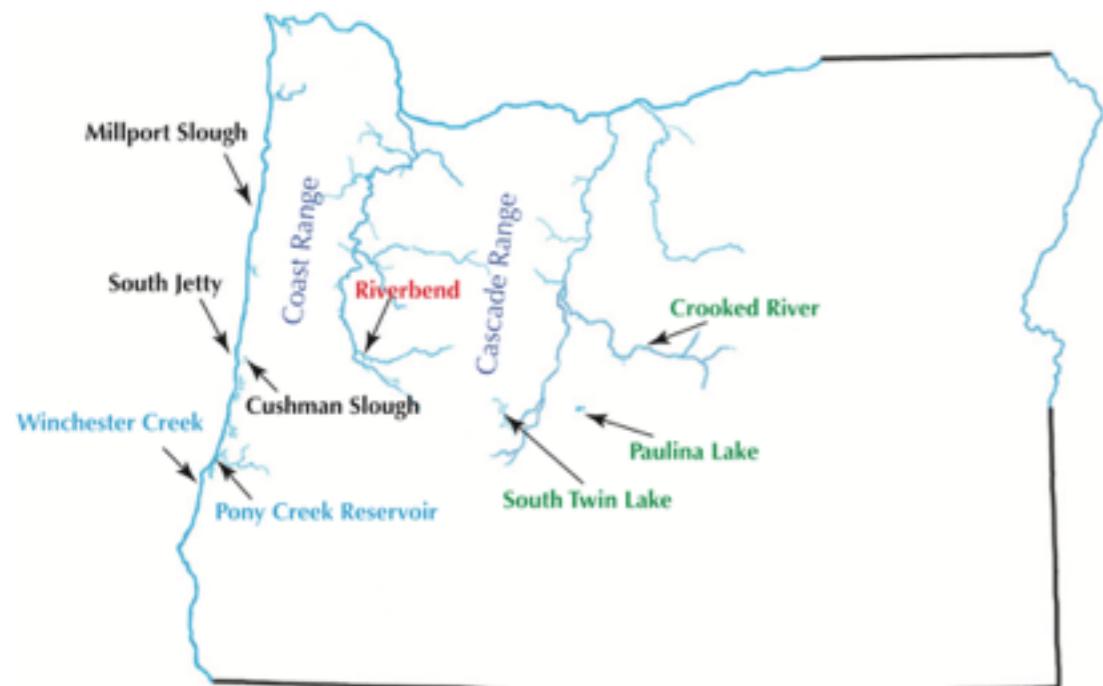


Parameters to control stack formation, ctd.

-n <distance between catalog loci>



Population Genetics without a Reference



The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing

JULIAN CATCHEN,¹ SUSAN BASSHAM,² TAYLOR WILSON, MARK CURREY, CONOR O'BRIEN,
QUICK YEATES and WILLIAM A. CRESKO
¹Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403, USA



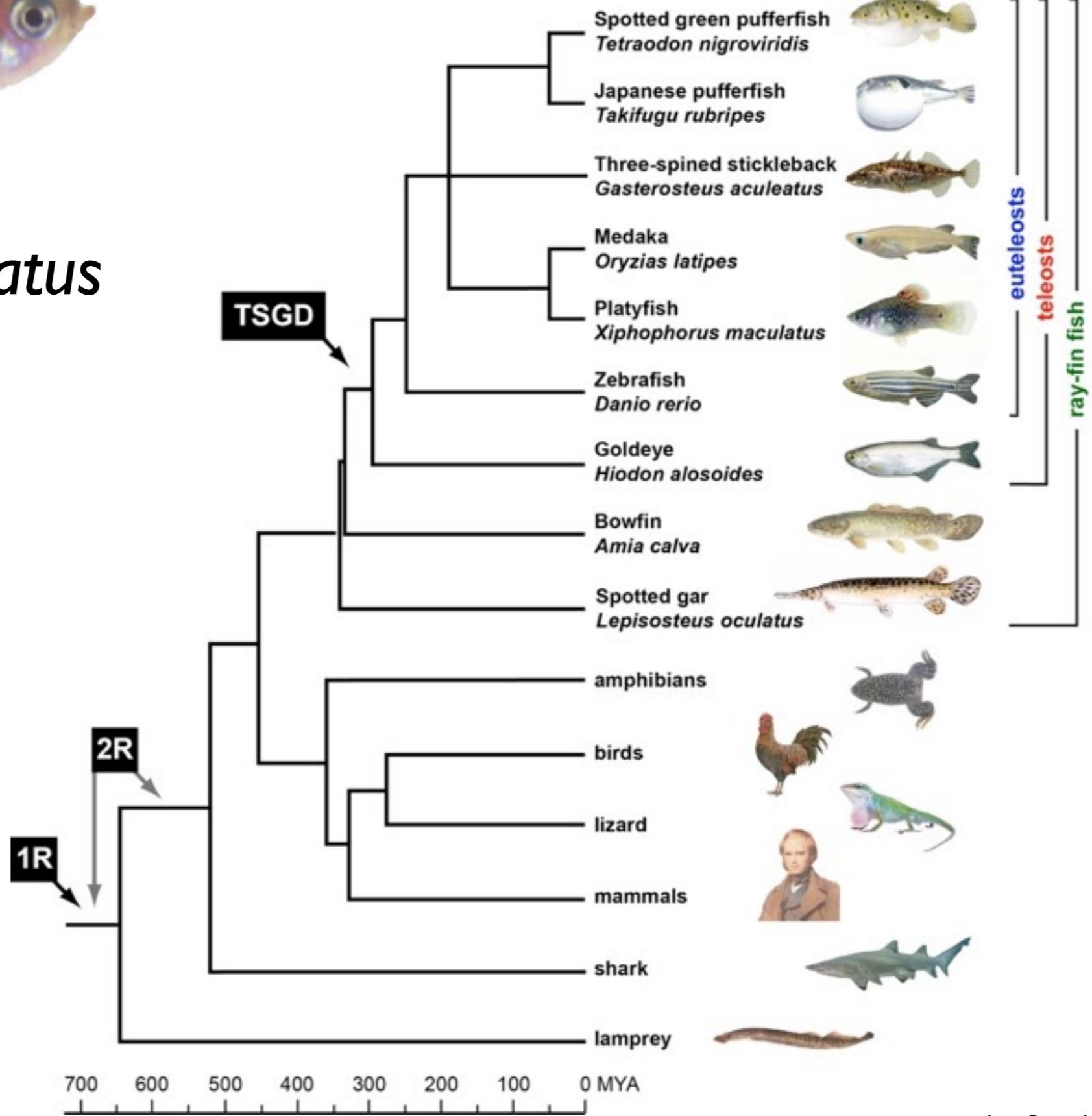
Xiphophorous maculatus



Angel
Amores



John
Postlethwait



Version 1.0

Copyright © 2004 by the Genetics Society of America
DOI: 10.1534/genetics.103.019349

A Microsatellite Genetic Linkage Map for *Xiphophorus*

R. B. Walter,^{*,†} J. D. Rains,^{*} J. E. Russell,^{*} T. M. Guerra,^{*} C. Daniels,^{*} Dennis A. Johnston,[†]
Jay Kumar,[†] A. Wheeler,^{*} K. Kelnar,^{*} V. A. Khanolkar,^{*} E. L. Williams,^{*} J. L. Hornecker,^{*}
L. Hollek,^{*} M. M. Mamerow,^{*} A. Pedroza^{*} and S. Kazianis^{*}

^{*}Molecular Biosciences Research Group, Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas 78666 and
[†]Science Park Research Division, MD Anderson Cancer Center, University of Texas, Smithville, Texas 78957

290 markers

Version 2.0

269 total fish

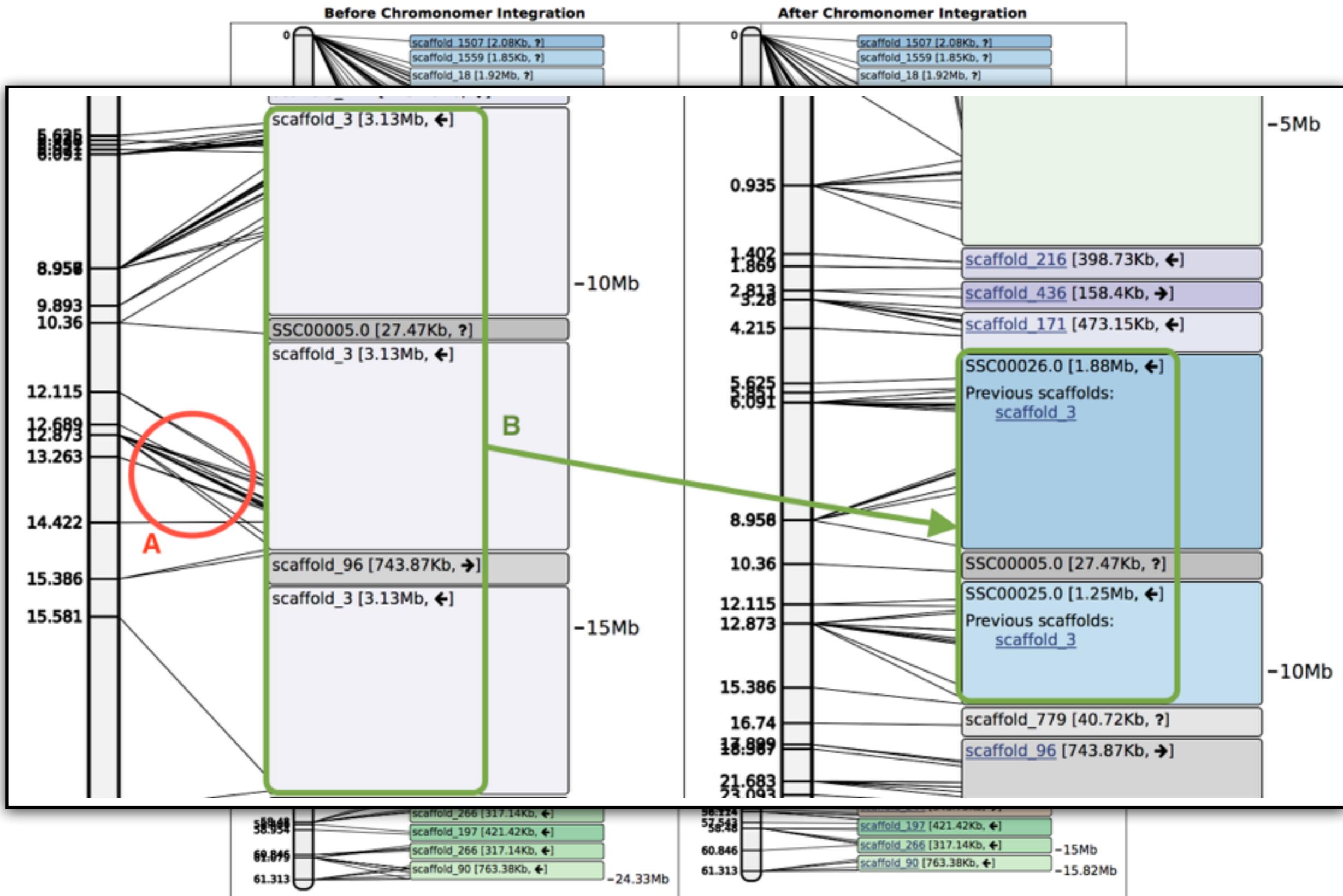
828,373,617 raw reads

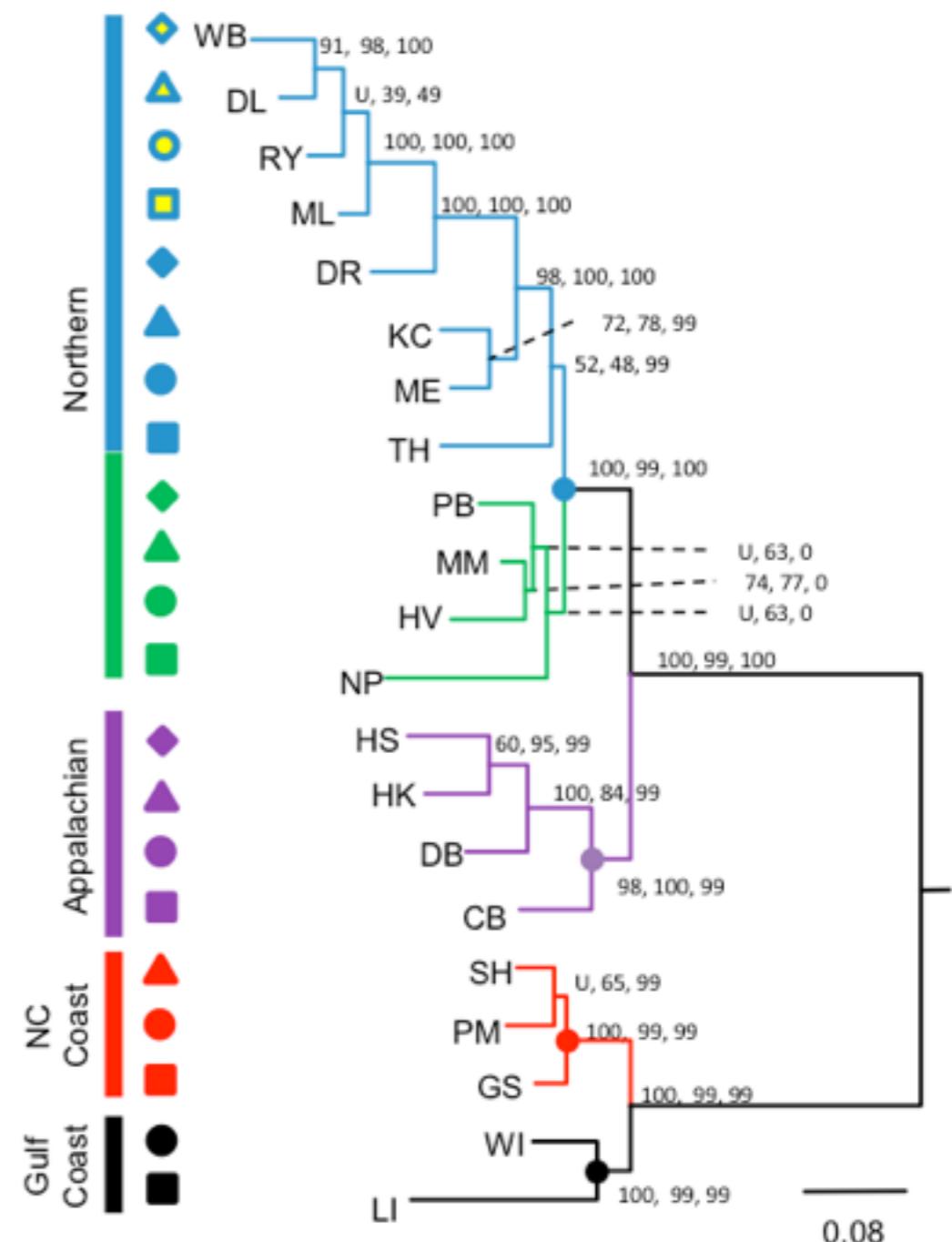
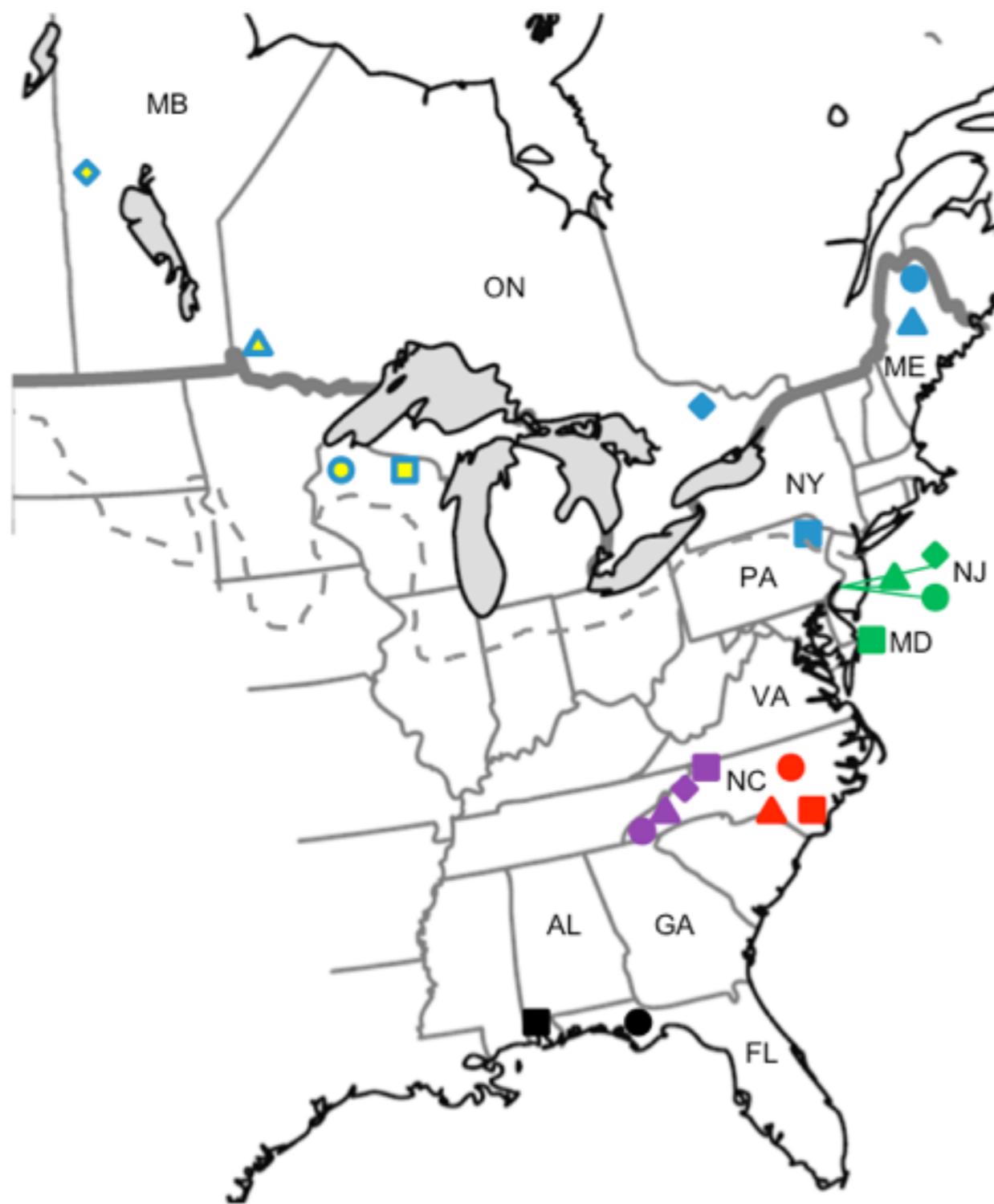
613,162,521 incorporated reads

22,144 markers in at least 90 progeny

Chromonomer

<http://catchenlab.life.illinois.edu/chromonomer/>

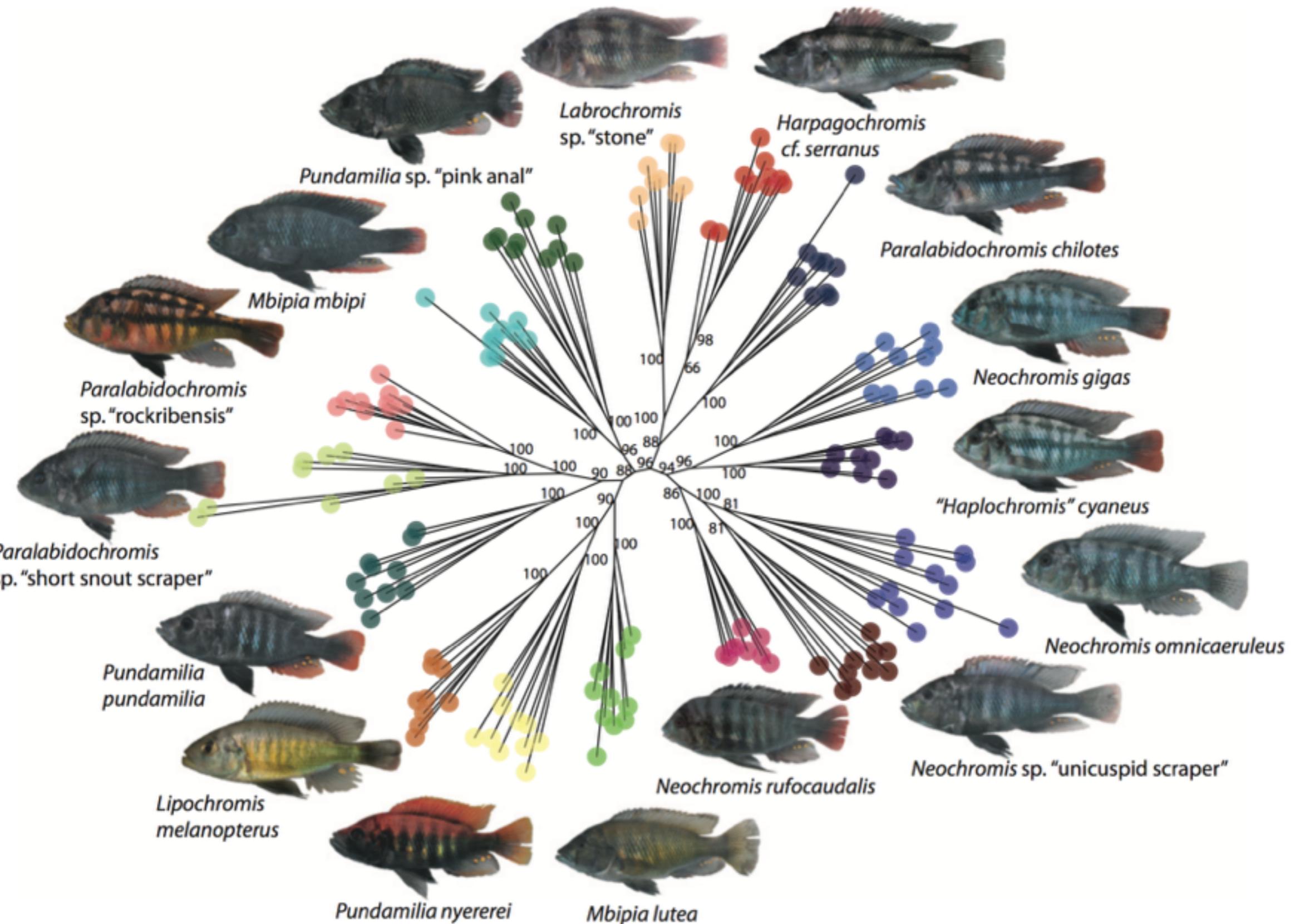




Resolving postglacial phylogeography using high-throughput sequencing

Kevin J. Emerson¹, Clayton R. Merz, Julian M. Catchen, Paul A. Hohenlohe, William A. Cresko, William E. Bradshaw, and Christina M. Holzapfel

¹Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, OR 97403-5289



Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation

CATHERINE E. WAGNER,^{*†} IRENE KELLER,^{*†} SAMUEL WITWER,^{*†} OLIVER M. SELZ,^{*}

SALOME MWAIKO,^{*†} LUCIE GREUTER,^{*†} ARJUN SIVASUNDAR^{*‡} and OLE SEEHAUSEN^{*†}

^{*}Department of Fish Ecology & Evolution, EAWAG Centre for Ecology, Evolution and Biogeochemistry, Seestrasse 79, 6547, Kastanienbaum, Switzerland, [†]Institute of Ecology and Evolution, Aquatic Ecology, University of Bern, Baltzerstrasse 6, 3012, Bern, Switzerland

RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping

**Omar A. Ali,* Sean M. O'Rourke,* Stephen J. Amish,[†] Mariah H. Meek,^{*,‡} Gordon Luikart,^{†,§}
Carson Jeffres,** and Michael R. Miller^{*,**,*}**

Sequence Capture of Dual-digest RADseq Libraries with Identifiable Duplicates (RADcap)

Sandra L. Hoffberg^{1,*}, Troy J. Kieran², Julian M. Catchen³, Alison Devault⁴, Brant C. Faircloth⁵, Rodney Mauricio¹, Travis C. Glenn^{1,2,*}