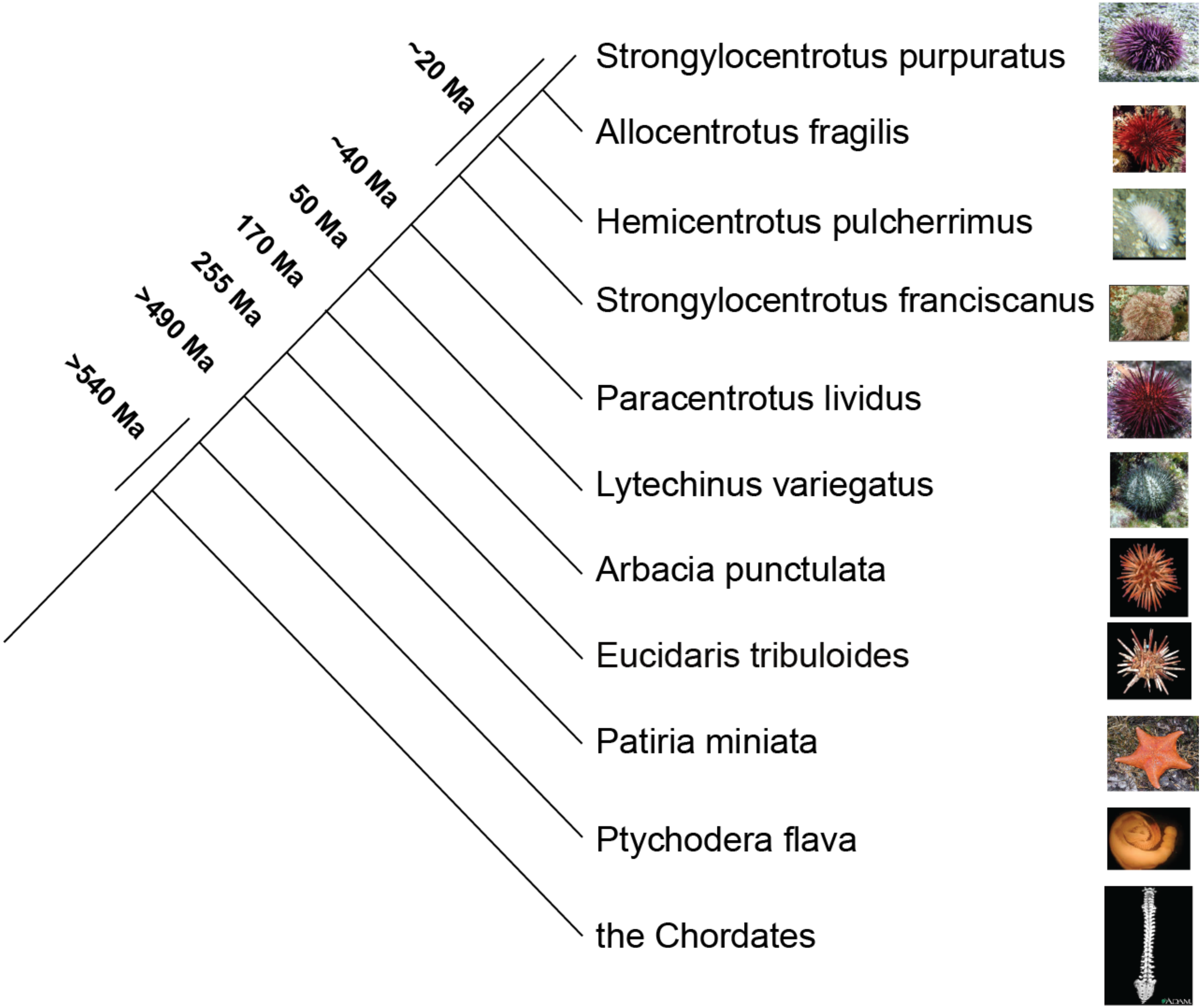
SEQUENCING ECHINODERM GENOMES

R. Andrew Cameron
David Felt
Ung-Jin Kim
Parul Kudtarkar
Manoj Samanta
Qiang Tu
Qiu (Autumn) Yuan

Center for Computational Regulatory Genomics
Beckman Institute
California Institute of Technology

Baylor College of Medicine Human Genome Sequencing Center



~20 Ma

~40 Ma

50 Ma

170 Ma

255 Ma

>490 Ma

>540 Ma

Strongylocentrotus purpuratus

Allocentrotus fragilis

Hemicentrotus pulcherrimus

Strongylocentrotus franciscanus

Paracentrotus lividus

Lytechinus variegatus

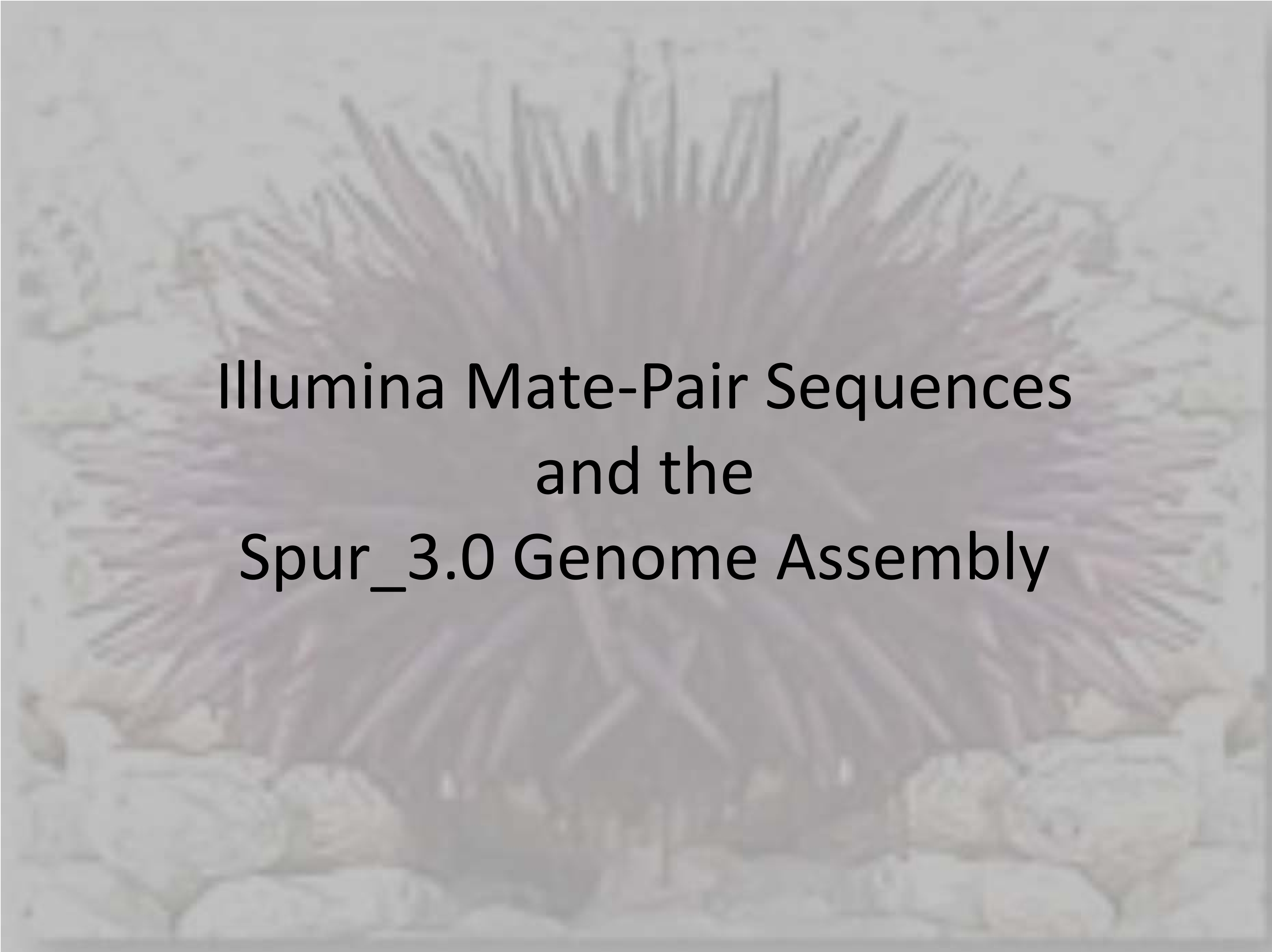
Arbacia punctulata

Eucidaris tribuloides

Patiria miniata

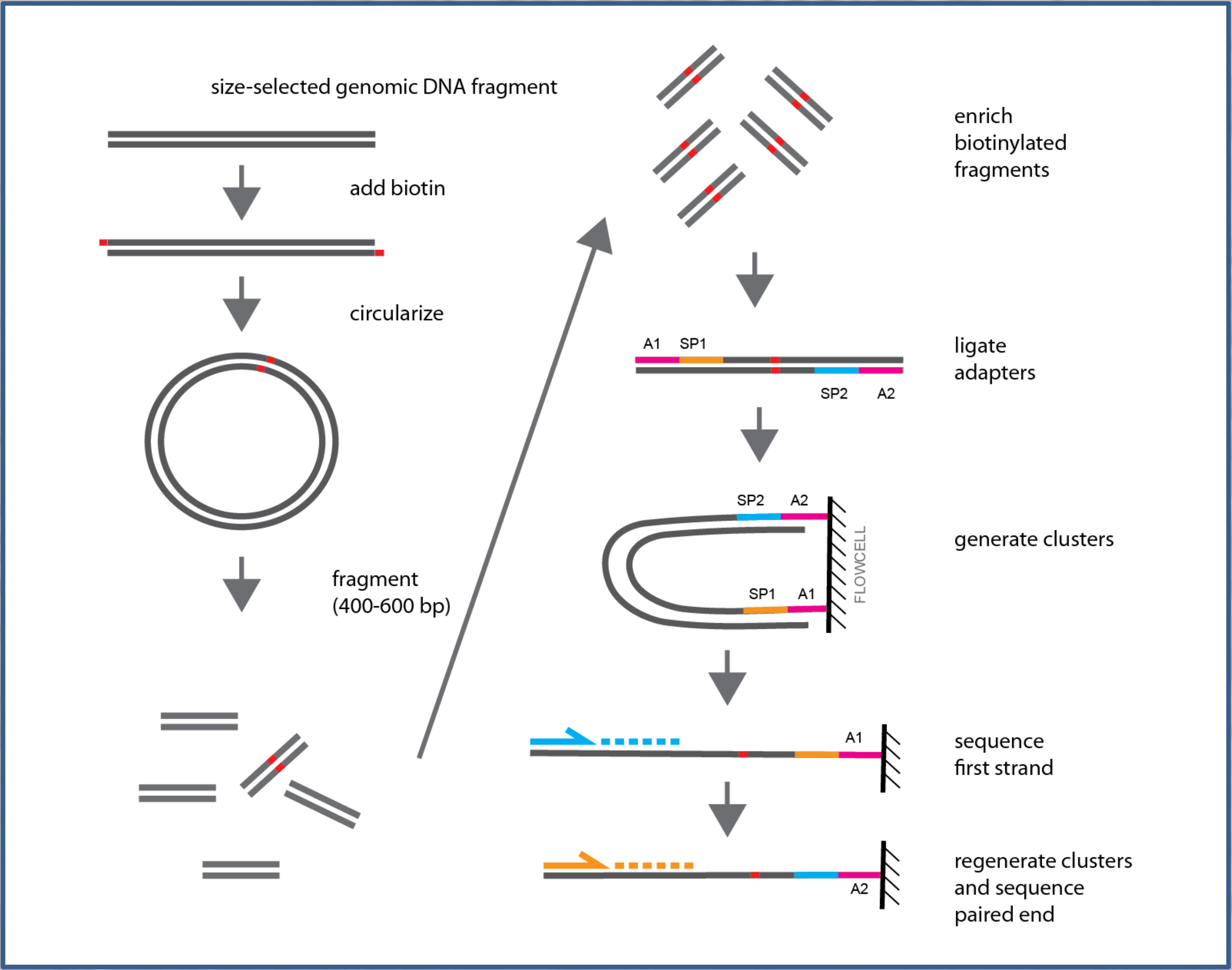
Ptychodera flava

the Chordates



Illumina Mate-Pair Sequences and the Spur_3.0 Genome Assembly

Illumina Paired-end Read Strategy



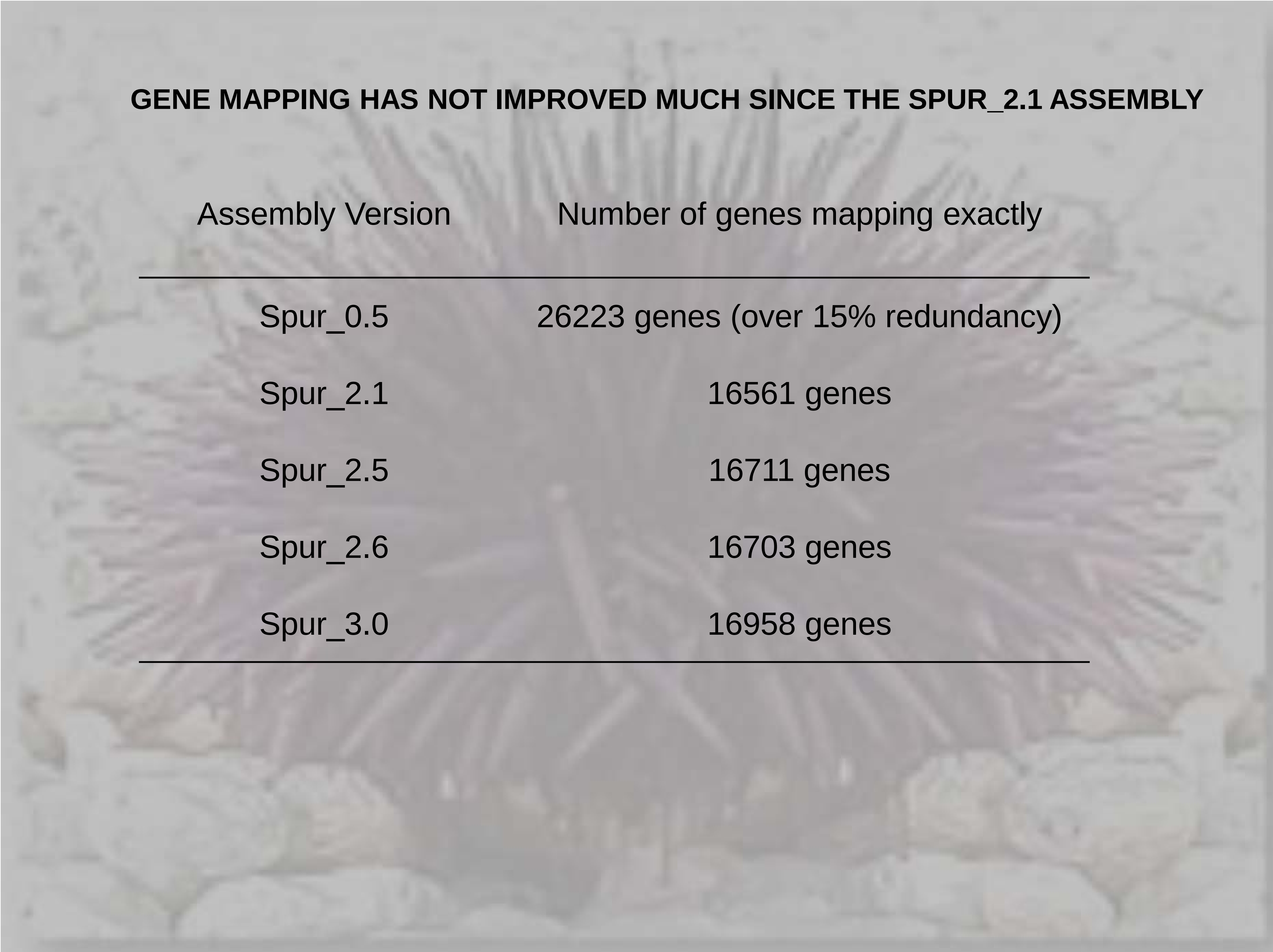
Assembly Statistics for the Purple Sea Urchin Genome

SCAFFOLDS (contigs oriented with paired end reads)

Assembly	Number	N50(kb)	Bases+Gaps(Mb)	Bases(Mb)
Spur_2.1	114,222	123,485	907Mb	810Mb
Spur_2.6	75,034	167,967	912Mb	806Mb
Spur_v3.0	32,008	401,883	935Mb	816Mb

CONTIGS (all sequence – no gaps)

Assembly	Number	N50(kb)	Bases+Gaps	Bases
Spur_2.6	196,827	11,503	806,143,440	806,143,440
Spur_v3.0	174,743	13,455	815,998,246	815,998,246



GENE MAPPING HAS NOT IMPROVED MUCH SINCE THE SPUR_2.1 ASSEMBLY

Assembly Version	Number of genes mapping exactly
Spur_0.5	26223 genes (over 15% redundancy)
Spur_2.1	16561 genes
Spur_2.5	16711 genes
Spur_2.6	16703 genes
Spur_3.0	16958 genes

0.5

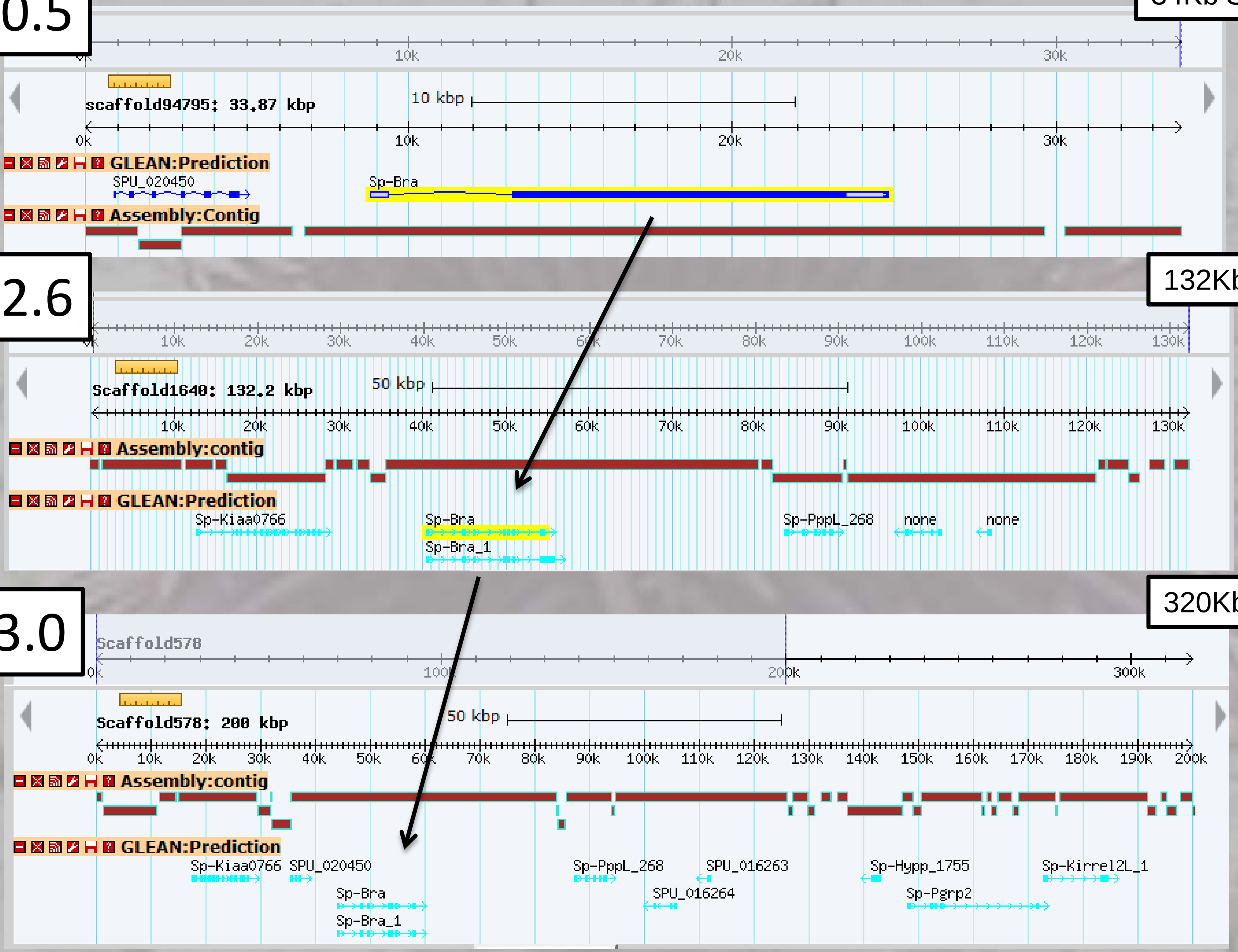
34Kb Scaffold

2.6

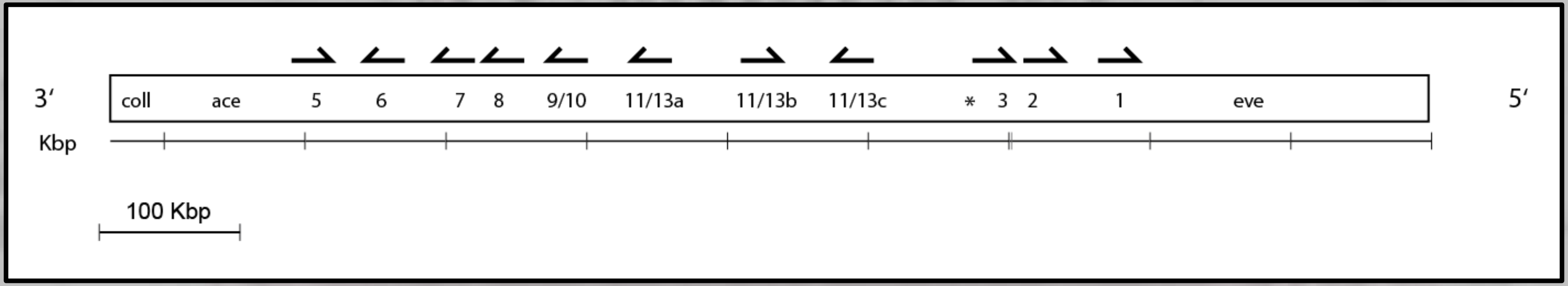
132Kb Scaffold

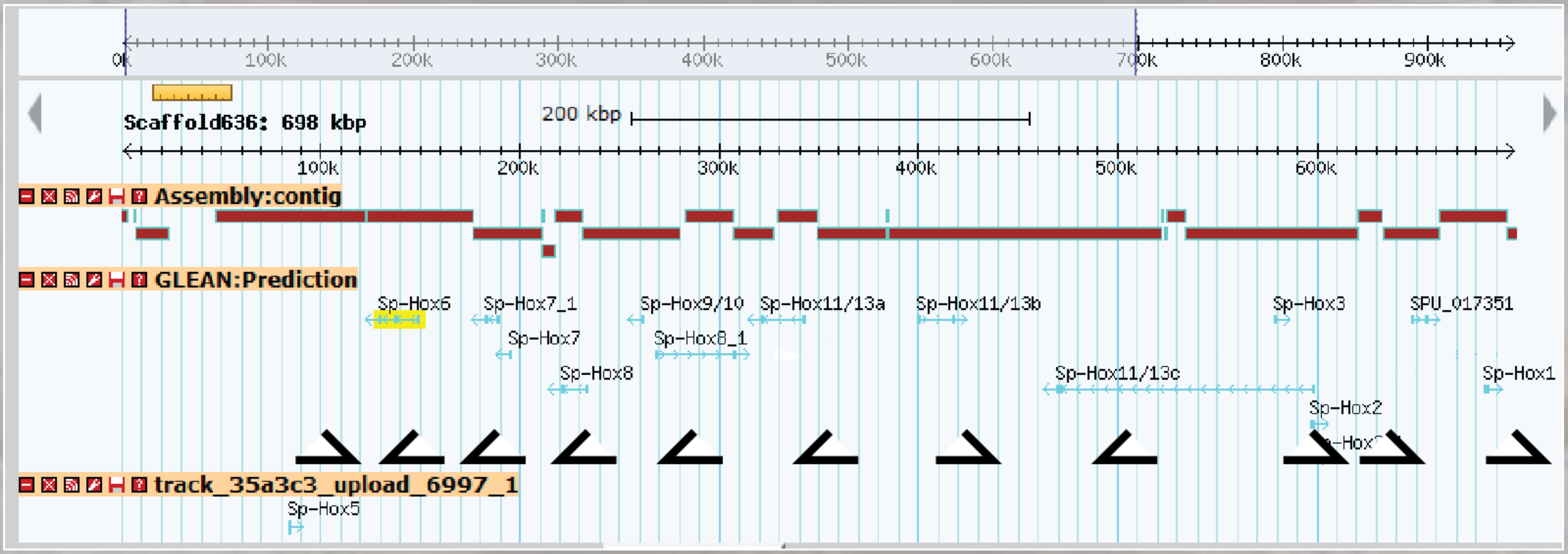
3.0

320Kb Scaffold

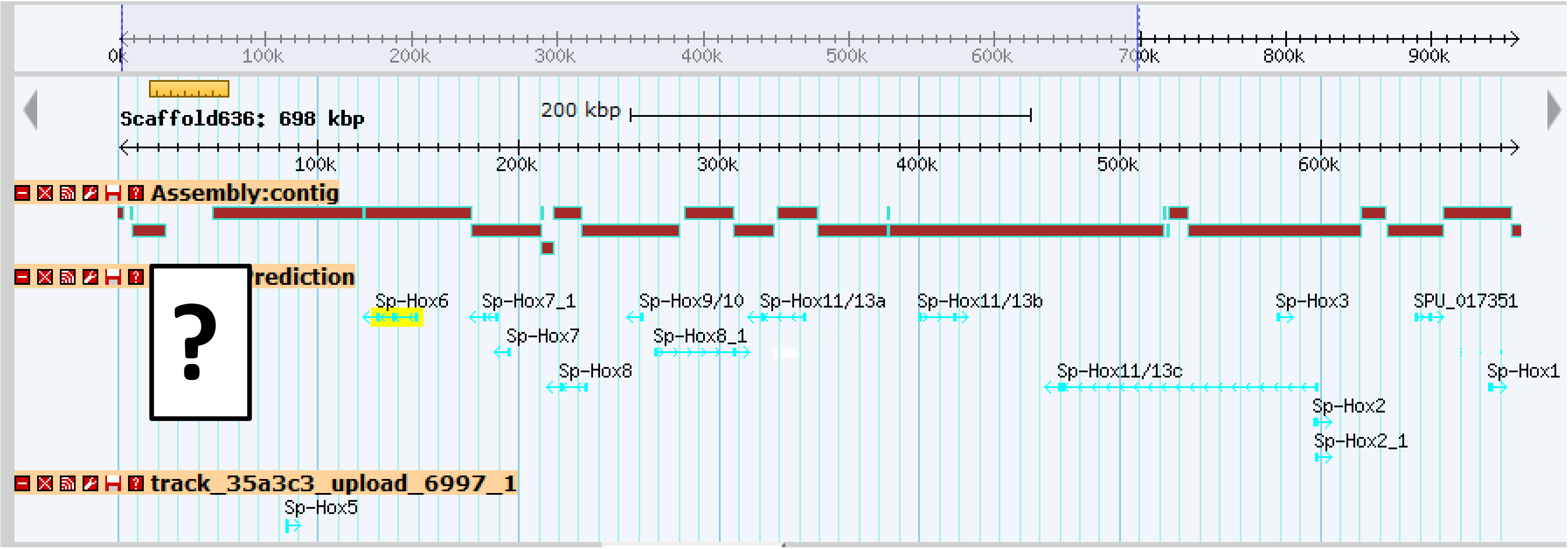


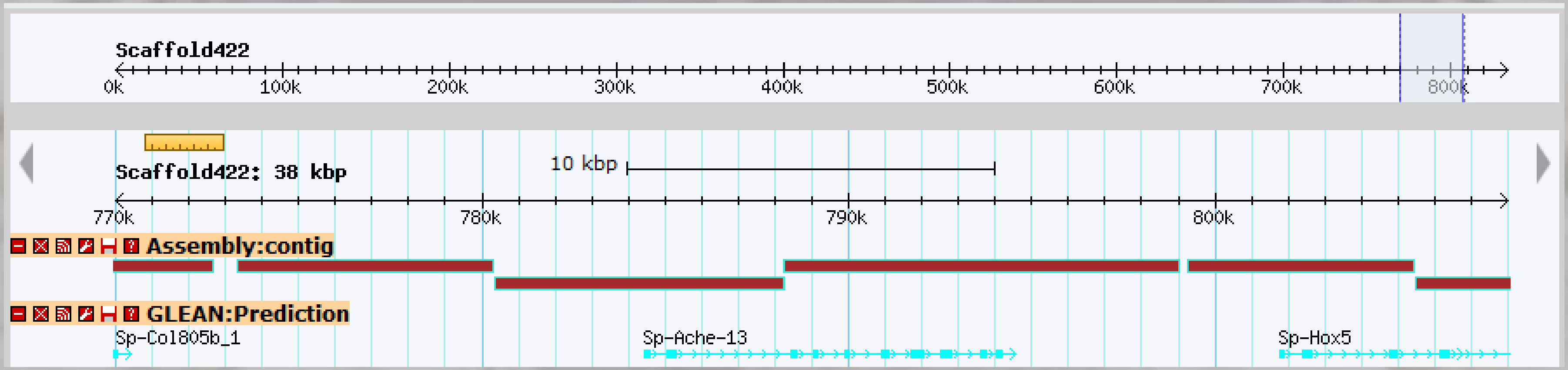
The Hox cluster is *almost* intact in the Spur 3.0 assembly

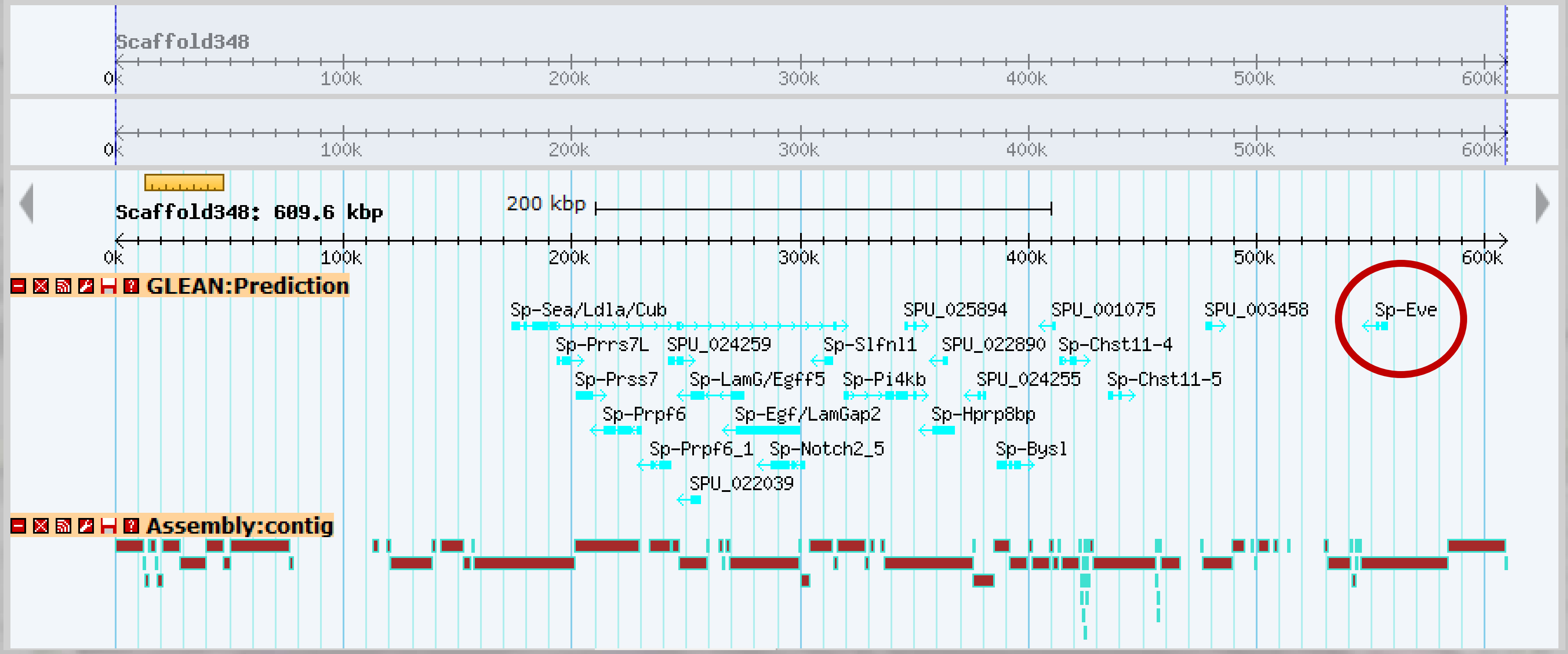




The Hox cluster external synteny remains in the Spur 3.0 assembly

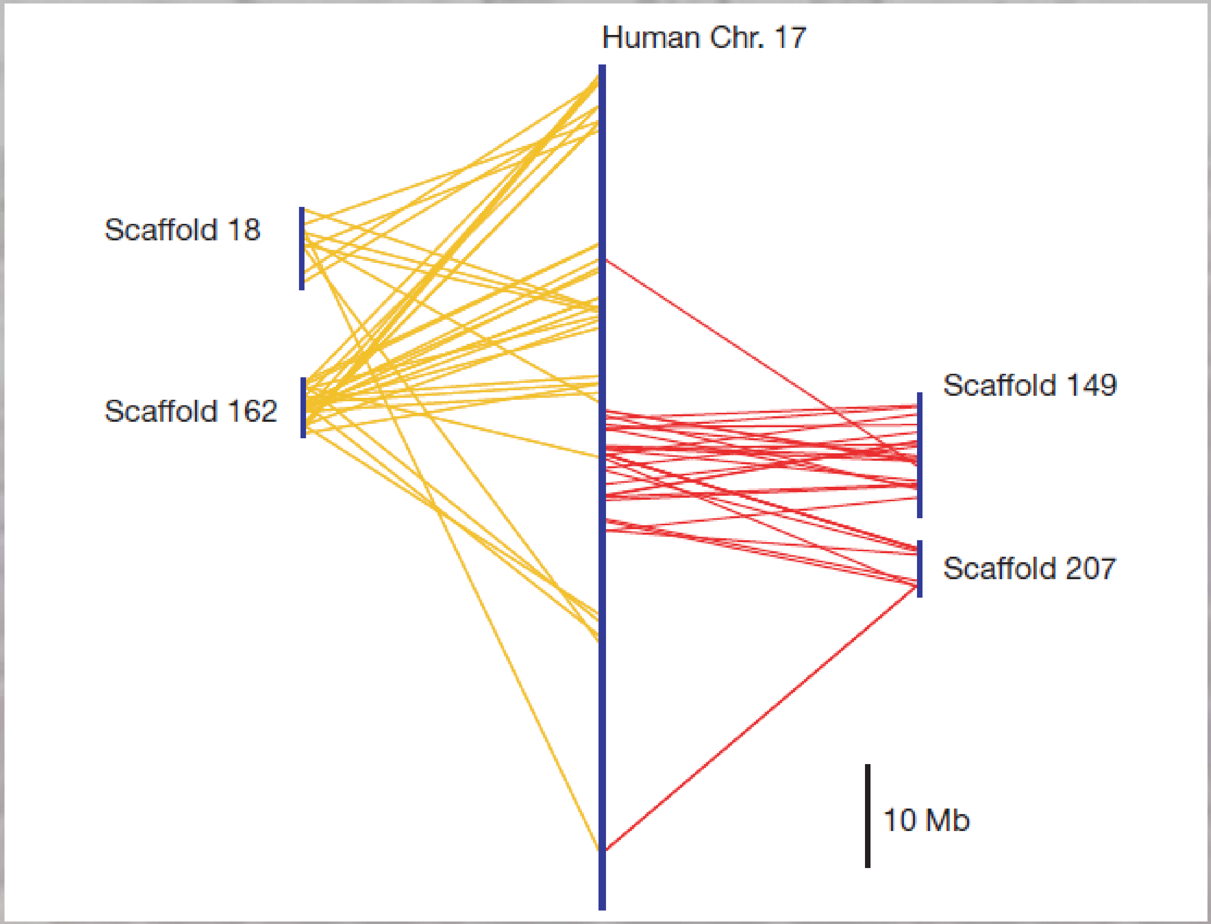






There is no overlap with the known synteny of Sp-Eve

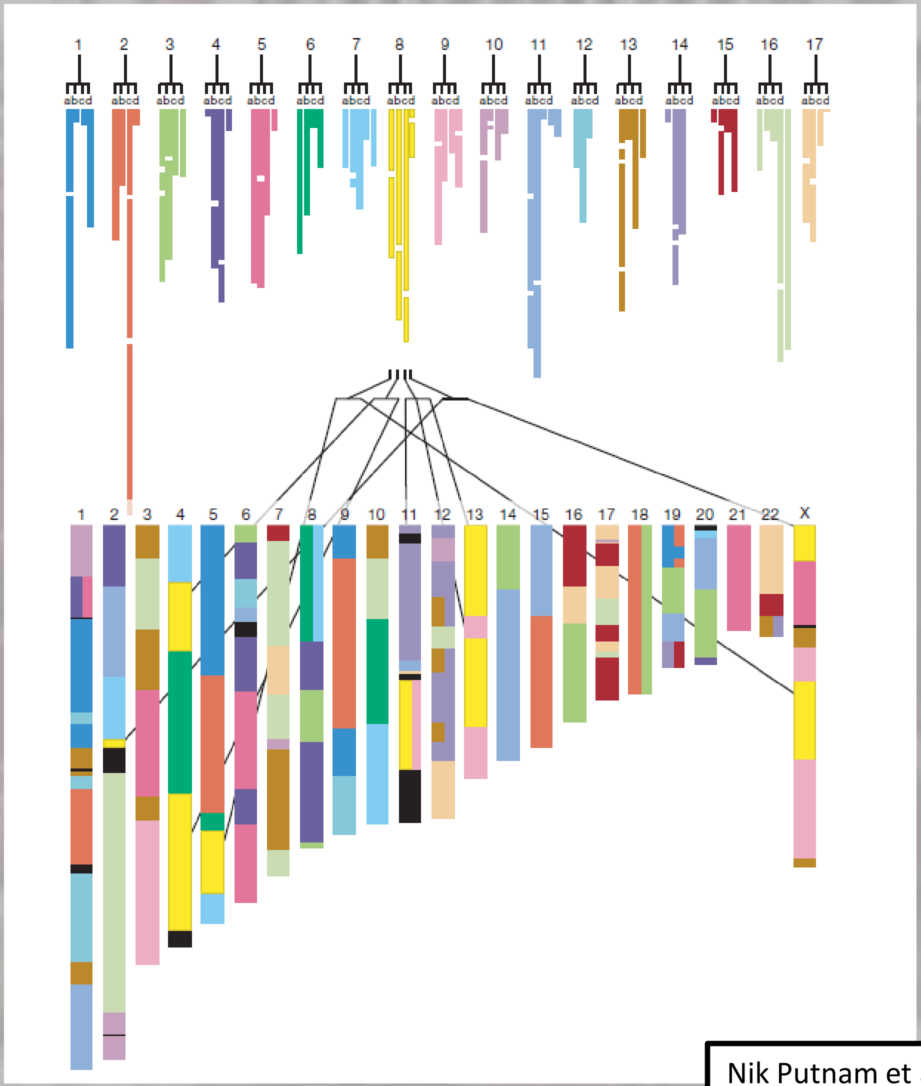
SYNTENY: SIMILAR BLOCKS OF GENES IN THE SAME RELATIVE POSITION IN THE GENOME.



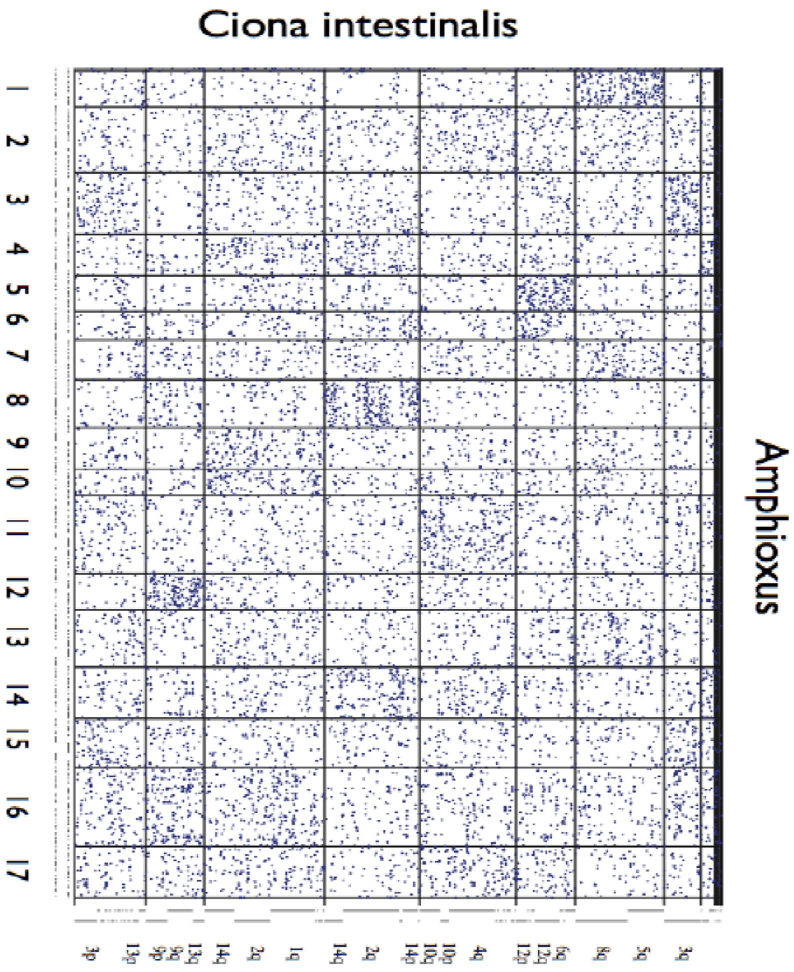
Orthologous genes from amphioxus scaffolds are concentrated in specific regions of the human chromosome 17.

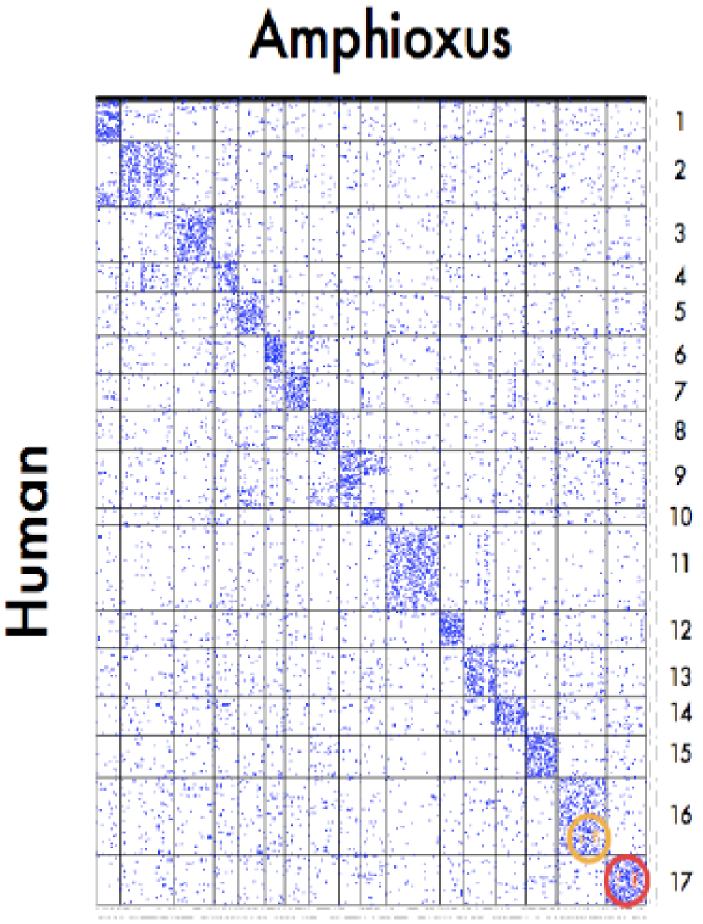
Nik Putnam et al. NATURE| Vol 453| 19 June 2008

Quadruple conserved synteny between human chromosome segments and Amphioxus scaffolds

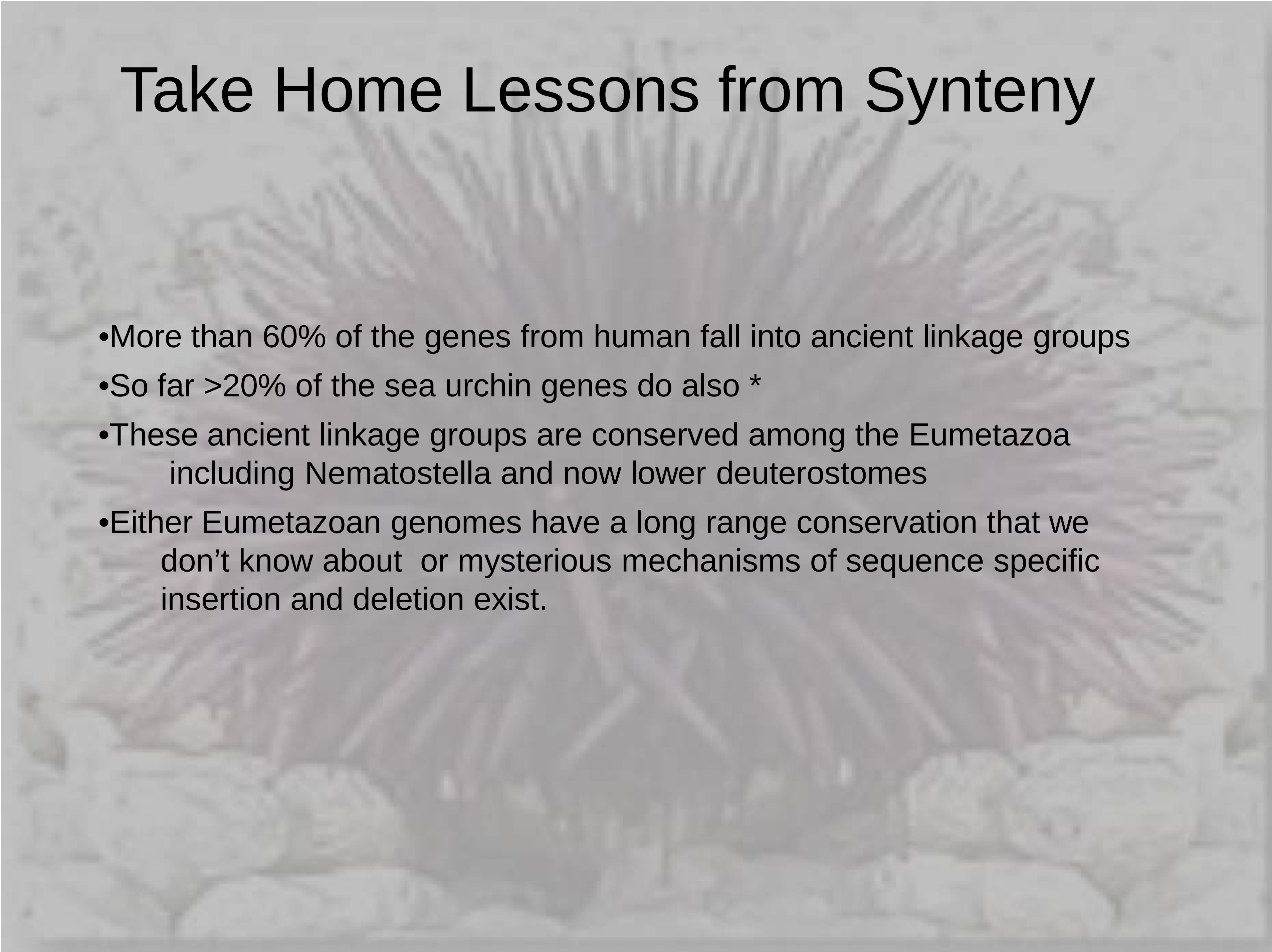


Nik Putnam et al. NATURE| Vol 453| 19 June 2008





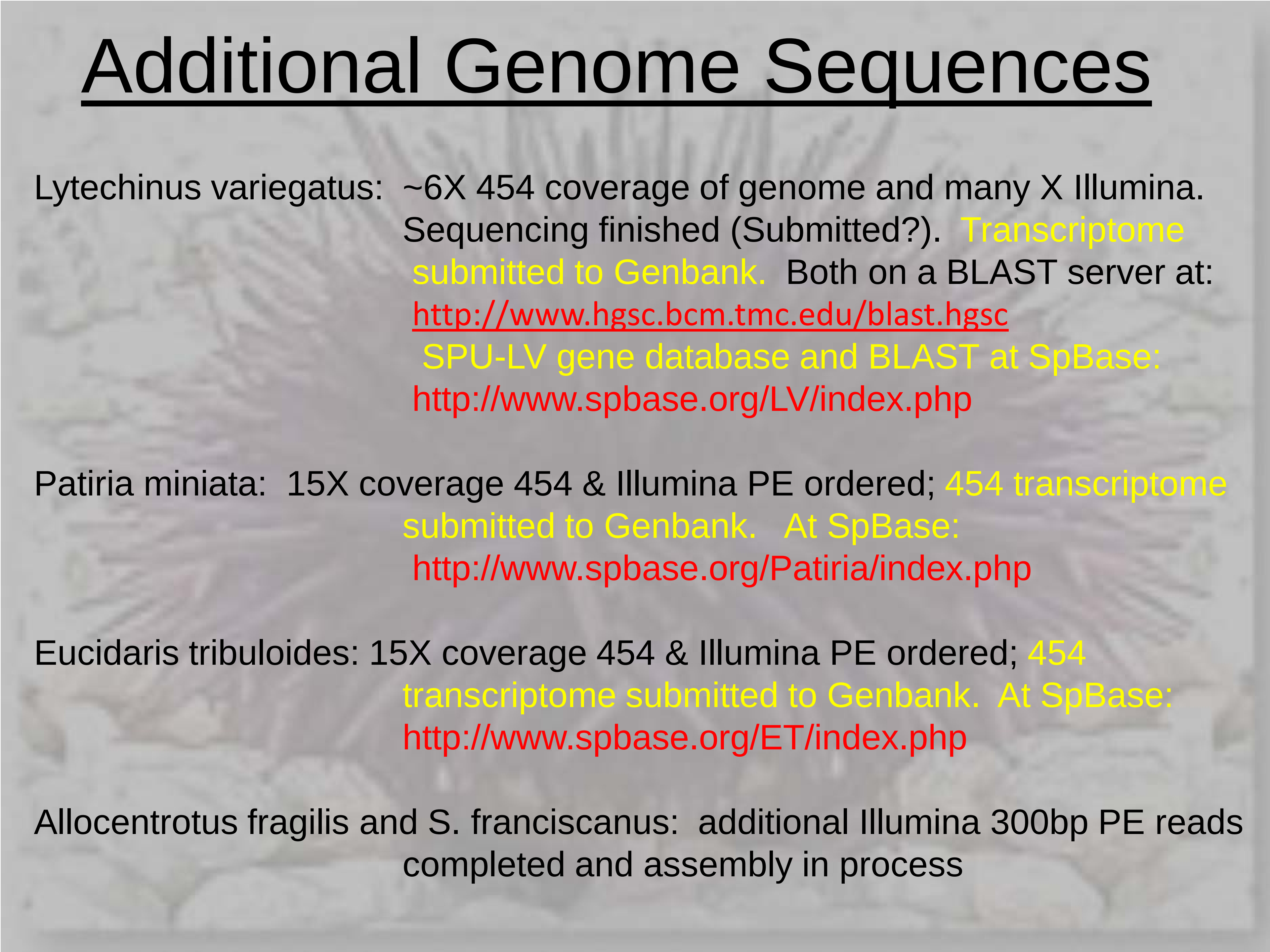
Nik Putnam et al. NATURE| Vol 453| 19 June 2008



Take Home Lessons from Synteny

- More than 60% of the genes from human fall into ancient linkage groups
- So far >20% of the sea urchin genes do also *
- These ancient linkage groups are conserved among the Eumetazoa including Nematostella and now lower deuterostomes
- Either Eumetazoan genomes have a long range conservation that we don’t know about or mysterious mechanisms of sequence specific insertion and deletion exist.

Additional Genome Sequences



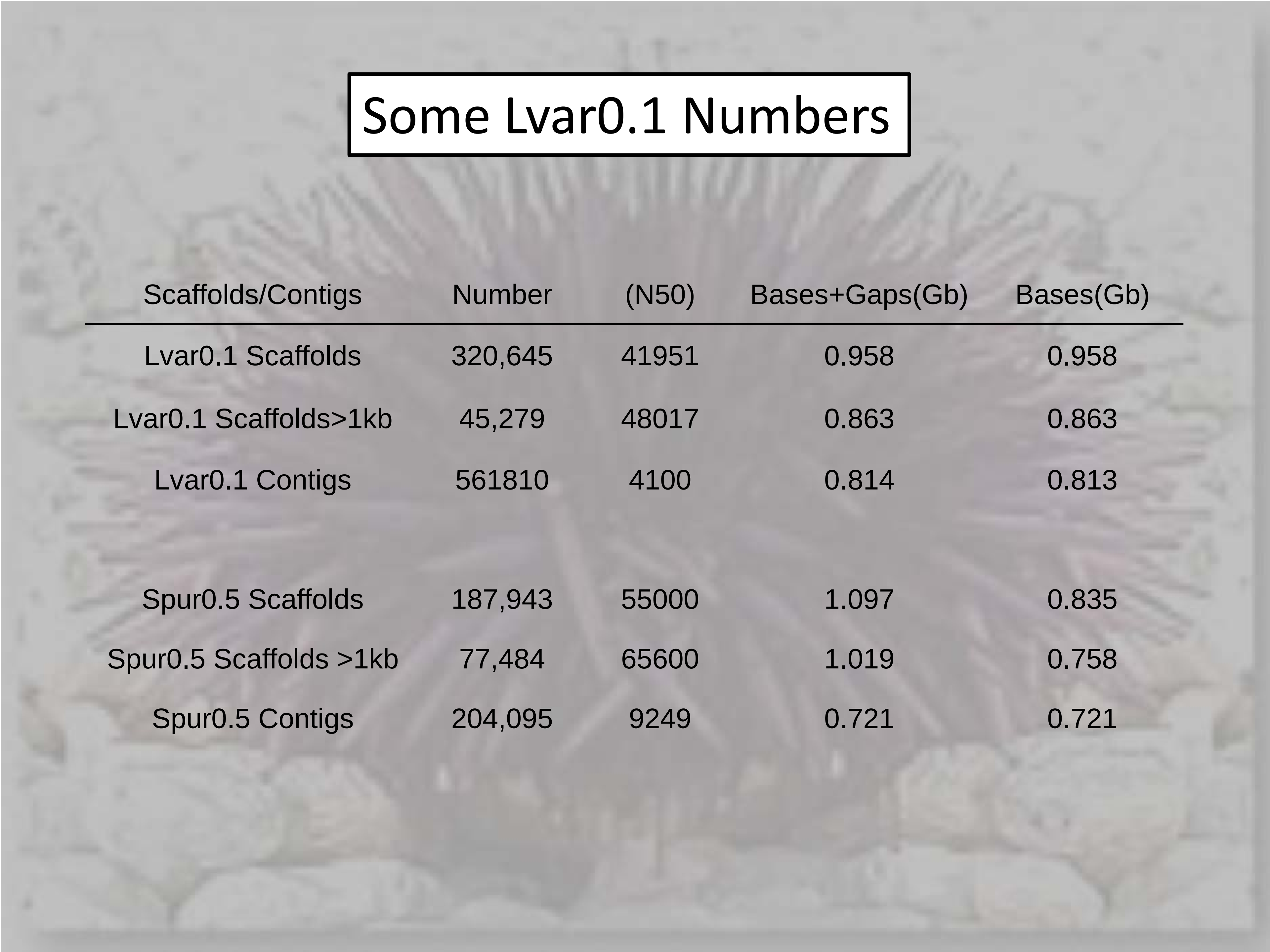
Lytechinus variegatus: ~6X 454 coverage of genome and many X Illumina. Sequencing finished (Submitted?). Transcriptome submitted to Genbank. Both on a BLAST server at:
http://www.hgsc.bcm.tmc.edu/blast.hgsc
SPU-LV gene database and BLAST at SpBase:
http://www.spbase.org/LV/index.php

Patiria miniata: 15X coverage 454 & Illumina PE ordered; 454 transcriptome submitted to Genbank. At SpBase:
http://www.spbase.org/Patiria/index.php

Eucidaris tribuloides: 15X coverage 454 & Illumina PE ordered; 454 transcriptome submitted to Genbank. At SpBase:
http://www.spbase.org/ET/index.php

Allocentrotus fragilis and S. franciscanus: additional Illumina 300bp PE reads completed and assembly in process

Some Lvar0.1 Numbers



Scaffolds/Contigs	Number	(N50)	Bases+Gaps(Gb)	Bases(Gb)
Lvar0.1 Scaffolds	320,645	41951	0.958	0.958
Lvar0.1 Scaffolds>1kb	45,279	48017	0.863	0.863
Lvar0.1 Contigs	561810	4100	0.814	0.813
Spur0.5 Scaffolds	187,943	55000	1.097	0.835
Spur0.5 Scaffolds >1kb	77,484	65600	1.019	0.758
Spur0.5 Contigs	204,095	9249	0.721	0.721

EchinoBase

Sequence information for various echinoderm species

Spbase is repository for small sequencing projects and a source of sequence information useful for gene discovery projects.

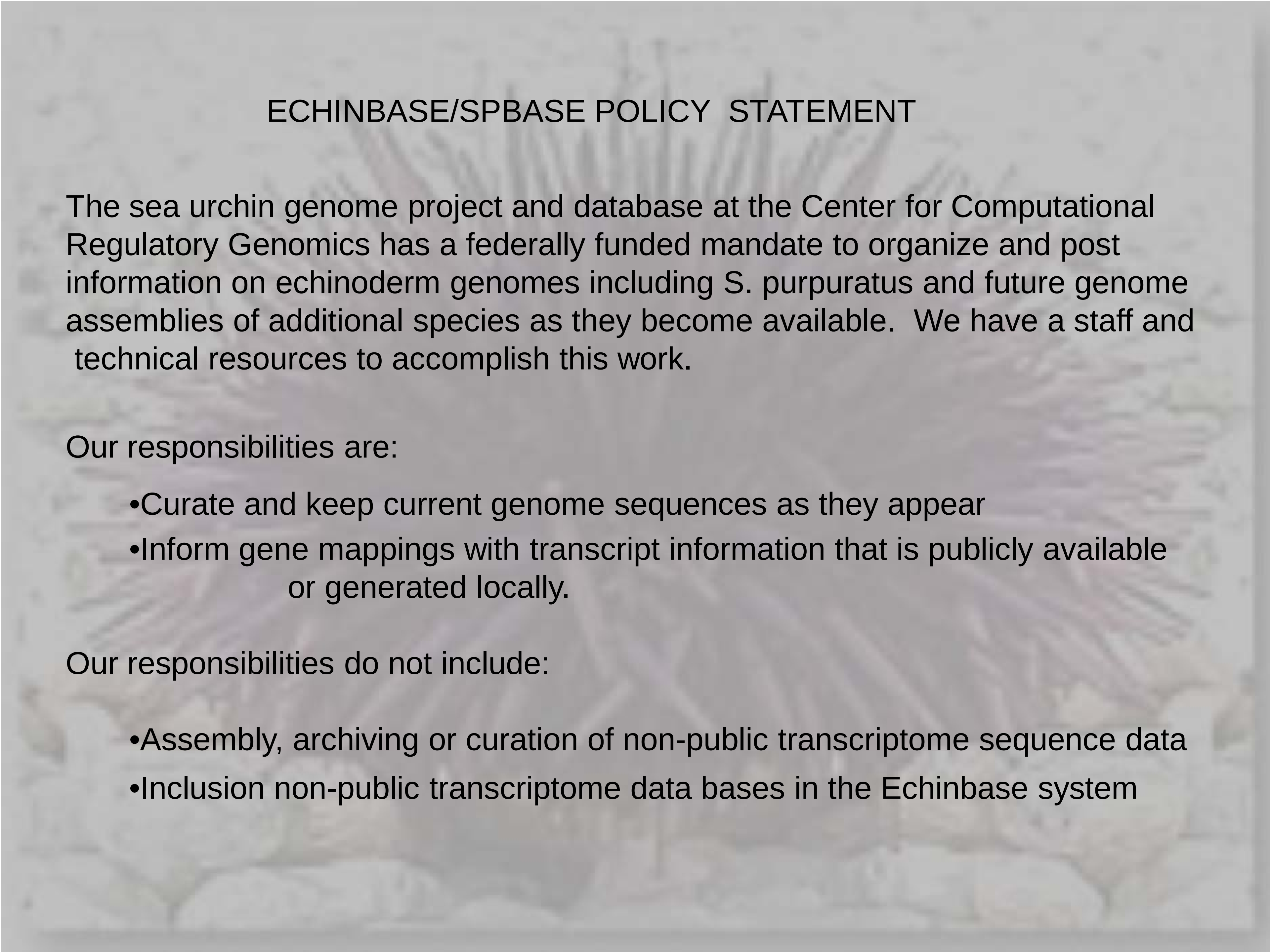
Access Strongylocentrotus purpuratus database and tools here

Welcome to the Echinoderm Genome Database

One of the particular advantages of the sea urchin as an experimental system lies in the well defined phylogeny that surrounds it. A deep and diverse fossil record and extensive biochemical studies together establish confident divergence times for many points in the phylogenetic tree for echinoderms (see figure). Sea urchins have diverged from the reference species at intervals covering 265 million years. Sea stars, hemichordates and the chordate branch to which humans belong diverged earlier. This character has been especially useful for the description of gene regulatory networks (GRN) and the cis-regulatory modules (CRM) of which they are made. One of the most comprehensive examples of a GRN sufficiently mature to demonstrate their full range of predictive and explanatory power are at present those worked out experimentally for the embryo of the sea urchin (Strongylocentrotus purpuratus, Sp). Genomic comparisons with another echinoid (sea urchin) species Lytechinus variegatus (Lv)

Echinoderm Database

- Strongylocentrotus purpuratus
- Patiria miniata
- Eucidaris tribuloides
- Allocentrotus fragilis
- Lytechinus variegatus



ECHINBASE/SPBASE POLICY STATEMENT

The sea urchin genome project and database at the Center for Computational Regulatory Genomics has a federally funded mandate to organize and post information on echinoderm genomes including S. purpuratus and future genome assemblies of additional species as they become available. We have a staff and technical resources to accomplish this work.

Our responsibilities are:

- Curate and keep current genome sequences as they appear
- Inform gene mappings with transcript information that is publicly available or generated locally.

Our responsibilities do not include:

- Assembly, archiving or curation of non-public transcriptome sequence data
- Inclusion non-public transcriptome data bases in the Echinbase system

ACKNOWLEDGEMENTS

BI Center for Computational Regulatory Genomics

Dave Felt
Ung-jin Kim
Parul Kudtarkar
Qiang Tu
Autumn Yuan
Manoj Samanta

Eric Davidson Lab Group

Eric Davidson
Patrick Leahy
Julie Hahn
Stefan Materna
Ping Dong
Feng Gao
Andy Ransick
Miki Yun
Julius Barsi
Jina Yun
Jongmin Nam
Isabelle Peter
Eric Erkenbrack
John Valencia
Enhu Li
Miao Cui
Erika Vielmas
Smadar ben Tabou De Leon

Baylor HGSC

Richard A. Gibbs
Kim Worley

Rice University

Nik Putnam
Paul Havlak

Support

NICHHD
NSF
NIH NCRR
Beckman Institute
NIGMS

ACKNOWLEDGEMENTS

BI Center for Computational Regulatory Genomics

Dave Felt Ung-jin Kim Parul Kudtarkar Qiang Tu
Autumn Yuan Manoj Samanta

Eric Davidson Lab Group

Eric Davidson Patrick Leahy Julie Hahn Stefan Materna
Ping Dong Feng Gao Andy Ransick Miki Yun
Julius Barsi Jina Yun Jongmin Nam Isabelle Peter
Eric Erkenbrack John Valencia Enhu Li Miao Cui
Erika Vielmas Smadar ben Tabou De Leon

Baylor HGSC

Richard A. Gibbs Kim Worley

Rice University

Nik Putnam Paul Havlak

Support

NICHHD NSF NIH NCRR Beckman Institute
NIGMS

END