Herpetology Evolution

Integrative Biology

Introduction to short read NGS:

Library construction, UCE capture and ddRADseq

The Natural History Museum, London

Autumn 2021

Instructor: Jeff Streicher

j.streicher@nhm.ac.uk

Litoria iris, Papua New Guinea

Unit 3: Targeted sequence capture of ultraconserved elements (UCEs)

https://github.com/nhm-herpetology/museum-NGS-training

Unit 2 Review

Lecture

- Inferring larger sequences from short read data
- Read mapping to reference sequences

Bioinformatics Lab

- How to assemble *de novo* contigs from reads
- How to map contigs/reads to references

Molecular Lab

- Shotgun Library Prep I
- Shotgun Library Prep II

Unit 3 Overview

Lecture

- Targeted Sequence Capture
- Ultraconserved elements and vertebrate phylogenetics

Bioinformatics Lab

- How to download and process UCE data

Molecular Lab

- Targeted sequence capture of UCEs from shotgun libraries made in Unit 2

Reduced-representation NGS sequencing

- Genomes can be large!
- We might want to compare multiple individuals/species
- Targeted Sequence Capture (TSC)
- Restriction site associated DNA sequencing (RADseq) [Unit 4]

Targeted sequence capture

- Hybridization-based capture
- Targeted enrichment
- NGS target enrichment
- Hybrid capture-based sequencing
- DNA bait capture
- Target capture

Targeted sequence capture

- Target enrichment is a cost-effective and efficient method for researchers to capture specific regions of interest after library preparation for NGS and has many advantages over whole genome sequencing (WGS). Target enrichment enables focused sequencing resources, which leads to reduced cost and simplified analysis.

Text from Roche.com

Sounds great… how does it work?

Published: 01 February 2009

Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing

Andreas Gnirke, Alexandre Melnikov, Jared Maguire, Peter Rogov, Emily M LeProust, William Brockman, Timothy Fennell, Georgia Giannoukos, Sheila Fisher, Carsten Russ, Stacey Gabriel, David B Jaffe, Eric S Lander & Chad Nusbaum

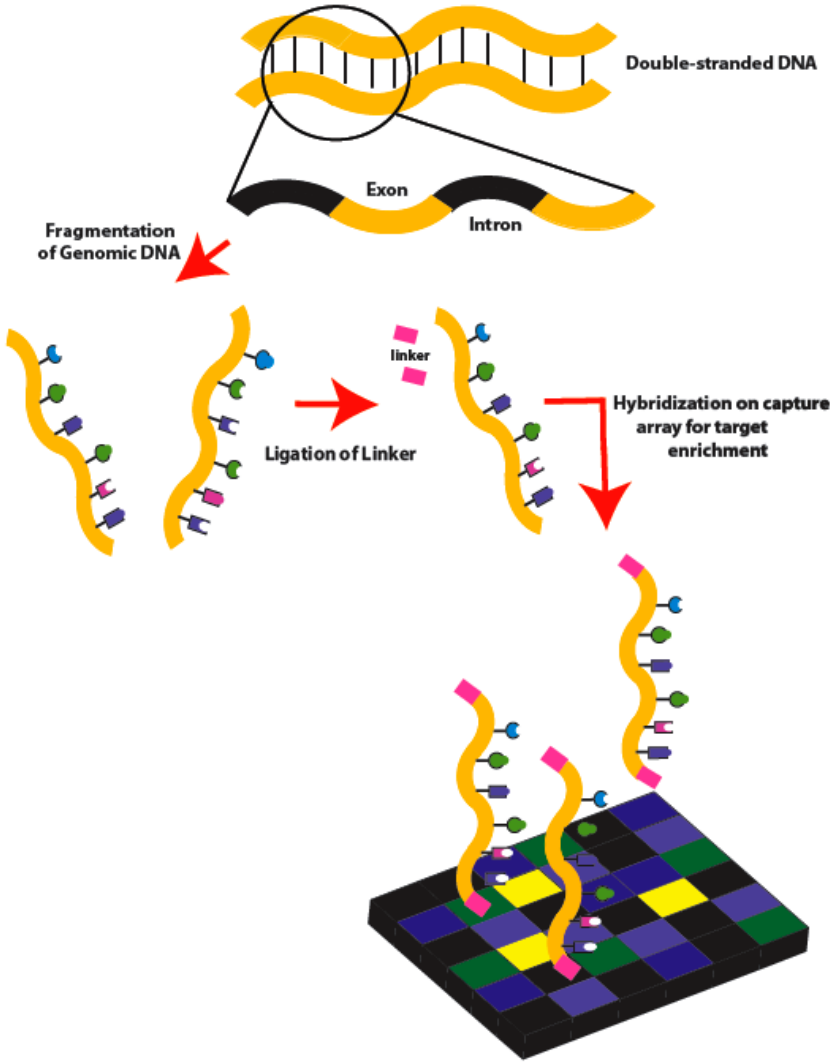
Nature Biotechnology **27**, 182–189 (2009) | Cite this article

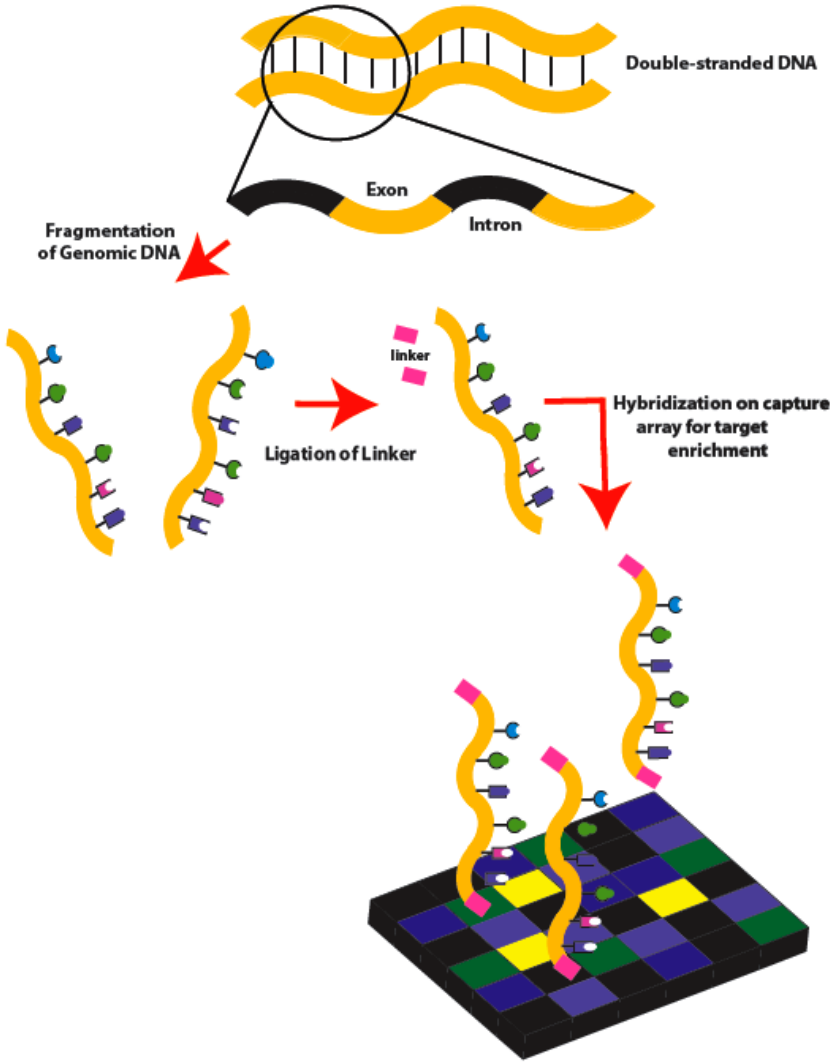
Abstract

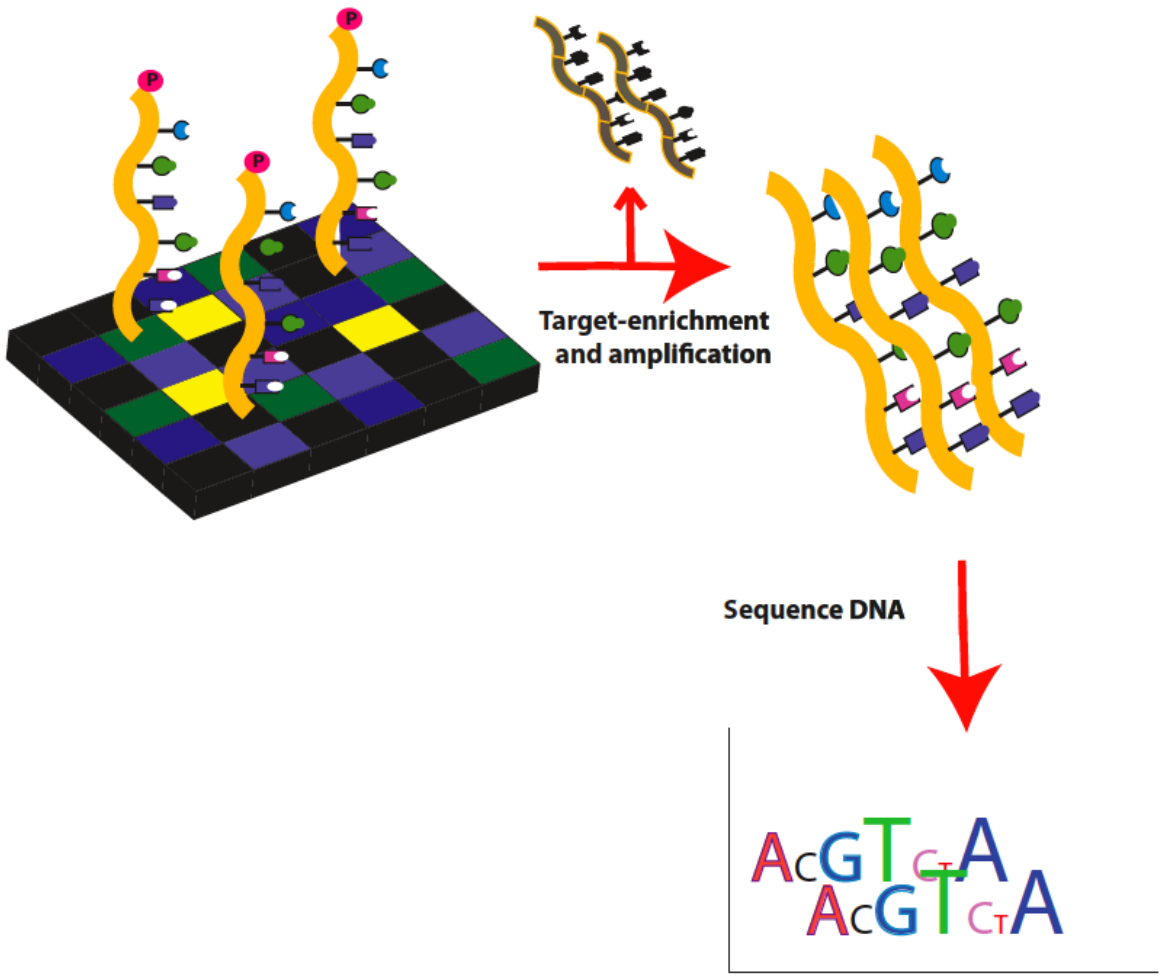
Targeting genomic loci by massively parallel sequencing requires new methods to enrich templates to be sequenced. We developed a capture method that uses biotinylated RNA 'baits' to fish targets out of a 'pond' of DNA fragments. The RNA is transcribed from PCR-amplified oligodeoxynucleotides originally synthesized on a microarray, generating sufficient bait for multiple captures at concentrations high enough to drive the hybridization. We tested this method with 170-mer baits that target >15,000 coding exons (2.5 Mb) and four regions (1.7 Mb total) using Illumina sequencing as read-out. About 90% of uniquely aligning bases fell on or near bait sequence; up to 50% lay on exons proper. The uniformity was such that ~60% of target bases in the exonic 'catch', and ~80% in the regional catch, had at least half the mean coverage. One lane of Illumina sequence was sufficient to call high-confidence genotypes for 89% of the targeted exon space.

Array-based sequence capture

- Microarrays contain single-stranded oligonucleotides with sequences to tile the region of interest fixed to the surface.

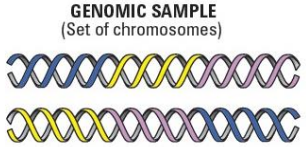






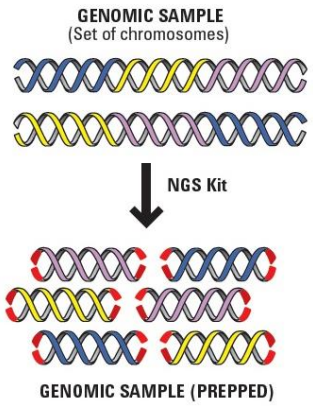
In-solution sequence capture

- ‘Free floating’ oligonucleotides hybridise to regions of interest and then are captured by magnetic beads.
- **UCEs are primarily enriched with this method**



Target Enrichment System Capture Process

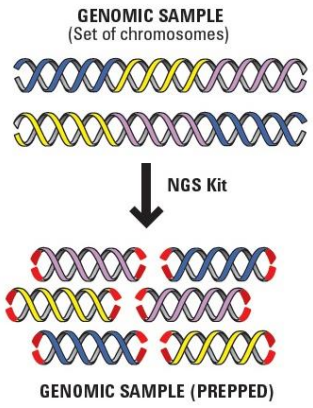
Image from Agilent.com



Target Enrichment System Capture Process

Image from Agilent.com

Unit 2 (Molecular Lab) →



Target Enrichment System Capture Process

Image from Agilent.com

Unit 2 (Molecular Lab) →

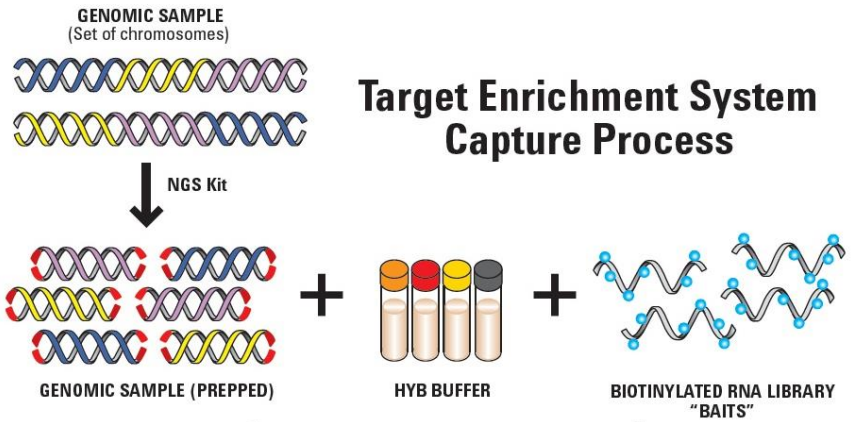
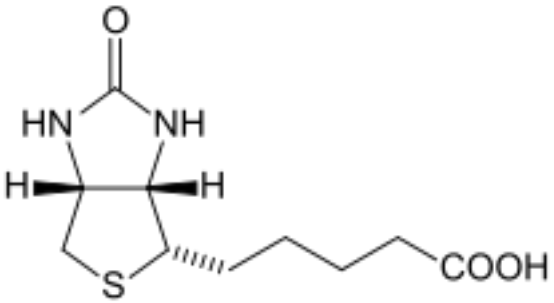


Image from Agilent.com

Biotinylation



- Biotin aka Vitamin B_7
- It covalently attaches to biological macromolecules (e.g. DNA/RNA and proteins)
- Biotinylation is rapid, specific and is unlikely to disturb the natural function of the molecule due to the small size of biotin (MW = 244.31 g/mol).

Text mostly from Wikipedia ☺

Unit 2 (Molecular Lab) →

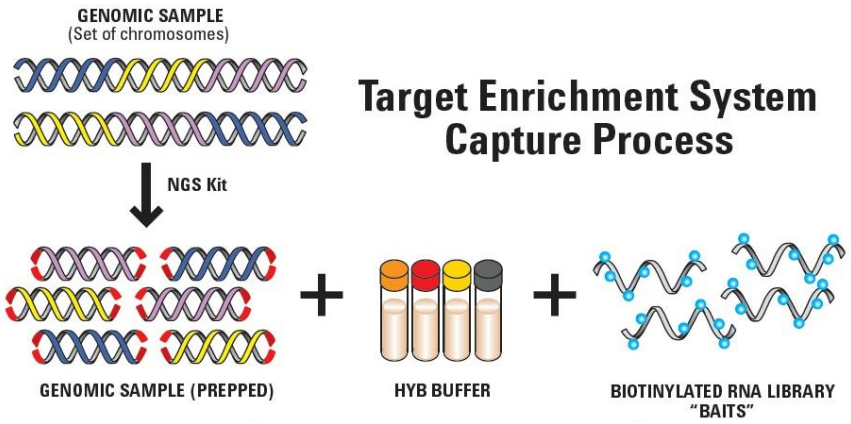


Image from Agilent.com

Unit 2 (Molecular Lab) →

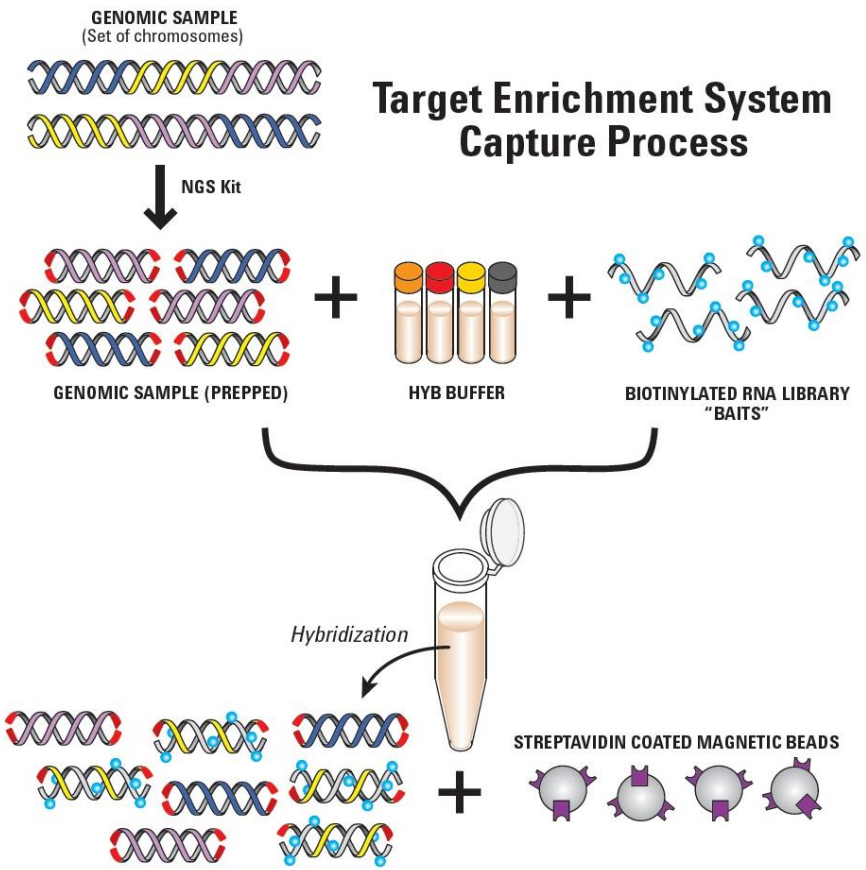


Image from Agilent.com

Streptavidin

- *Streptomyces avidinii*
- The strong streptavidin-biotin interaction can be used to attach various biomolecules to one another or onto a solid support.
- The magnetic beads for separation of biotinylated biomolecules have a Streptavidin ligand.

Streptomyces avidinii Type strain MA-833
Image by BacDive-DSMZ

Streptavidin-coated magnetic beads

- Streptavidin Magnetic Beads are **1 µm superparamagnetic particles** covalently coupled to a highly pure form of streptavidin. The beads can be used to capture biotin labeled substrates including antigens, antibodies and nucleic acids.
- Solid Phase Reversible Immobilization (SPRI) technology

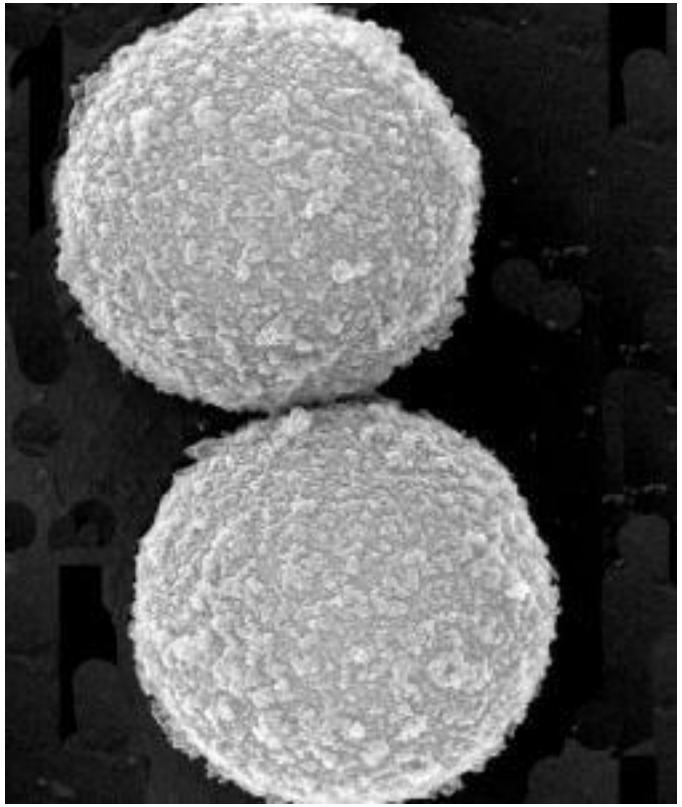
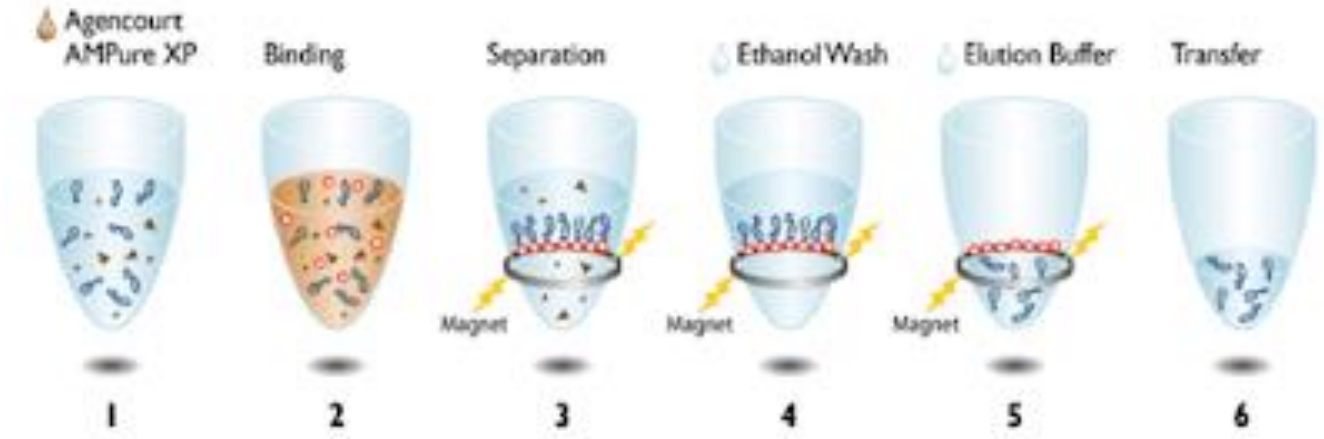


Image by VWR

An aside about SPRI beads…



Solid Phase Reversible Immobilization (SPRI) technology

- Sera-Mag Magnetic Speedbeads
- Not streptavidin-coated
- SPRI beads are paramagnetic (magnetic only in a magnetic field) and this prevents them from clumping and falling out of solution.
- Each bead is made of polystyrene surrounded by a layer of magnetite, which is coated with carboxyl molecules.

Image by Merck

Streptavidin-coated magnetic beads

- Streptavidin Magnetic Beads are **1 µm superparamagnetic particles** covalently coupled to a highly pure form of streptavidin. The beads can be used to capture biotin labeled substrates including antigens, antibodies and nucleic acids.
- Solid Phase Reversible Immobilization (SPRI) technology

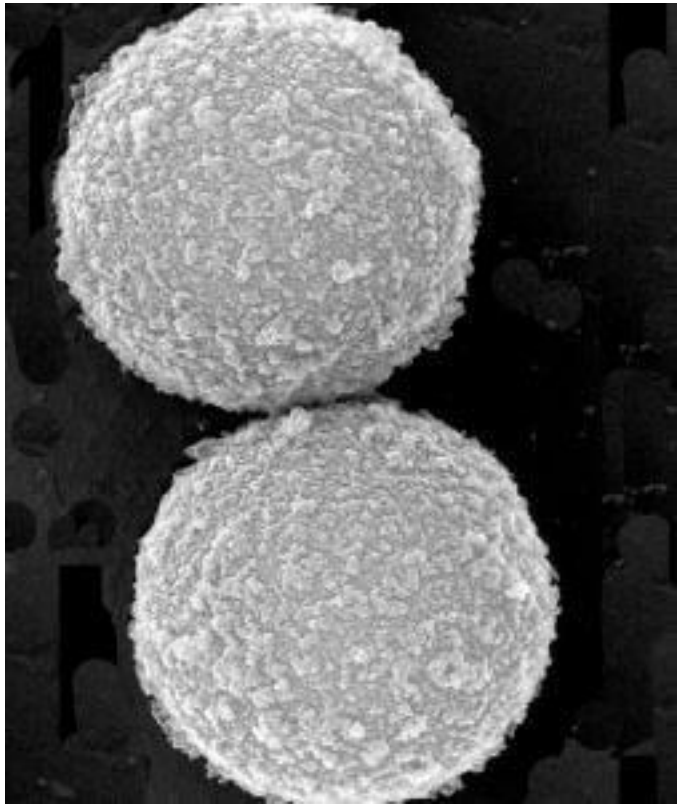


Image by VWR

Unit 2 (Molecular Lab) →

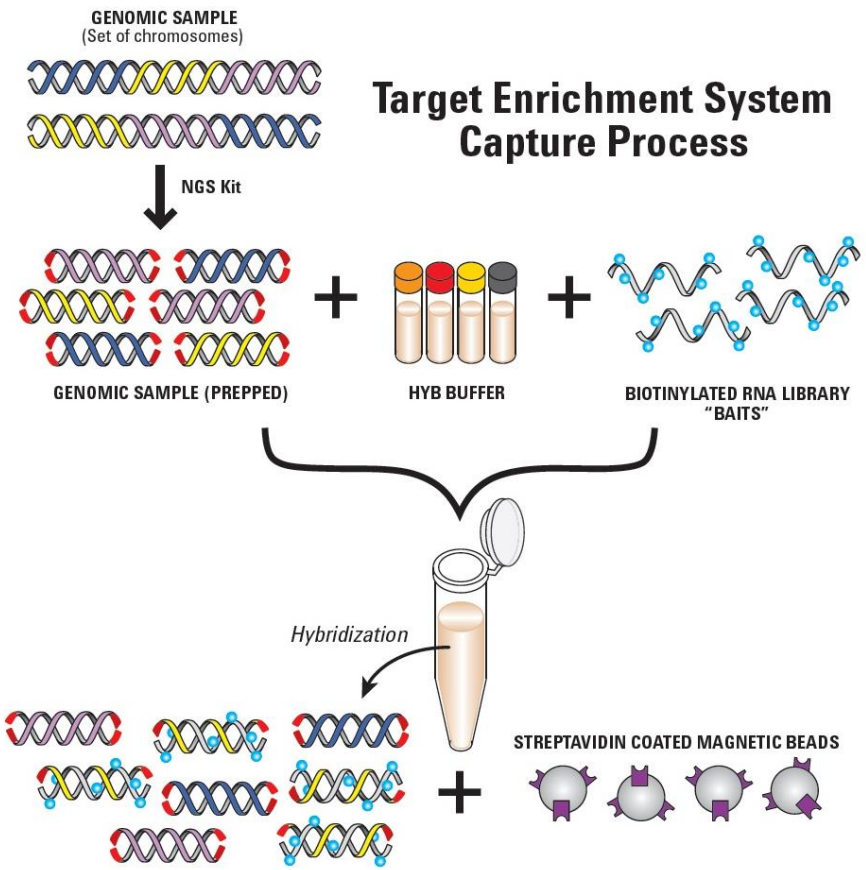


Image from Agilent.com

Unit 2 (Molecular Lab) →

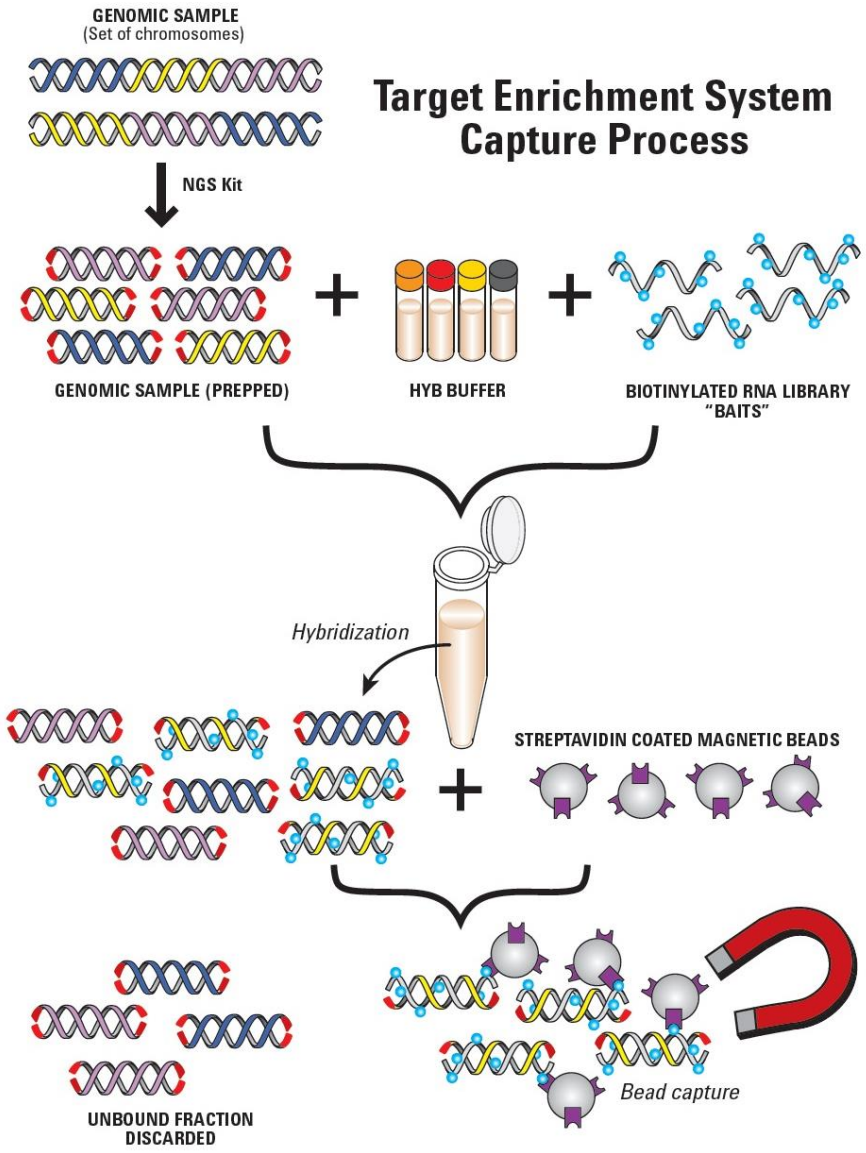


Image from Agilent.com

Unit 2 (Molecular Lab) →

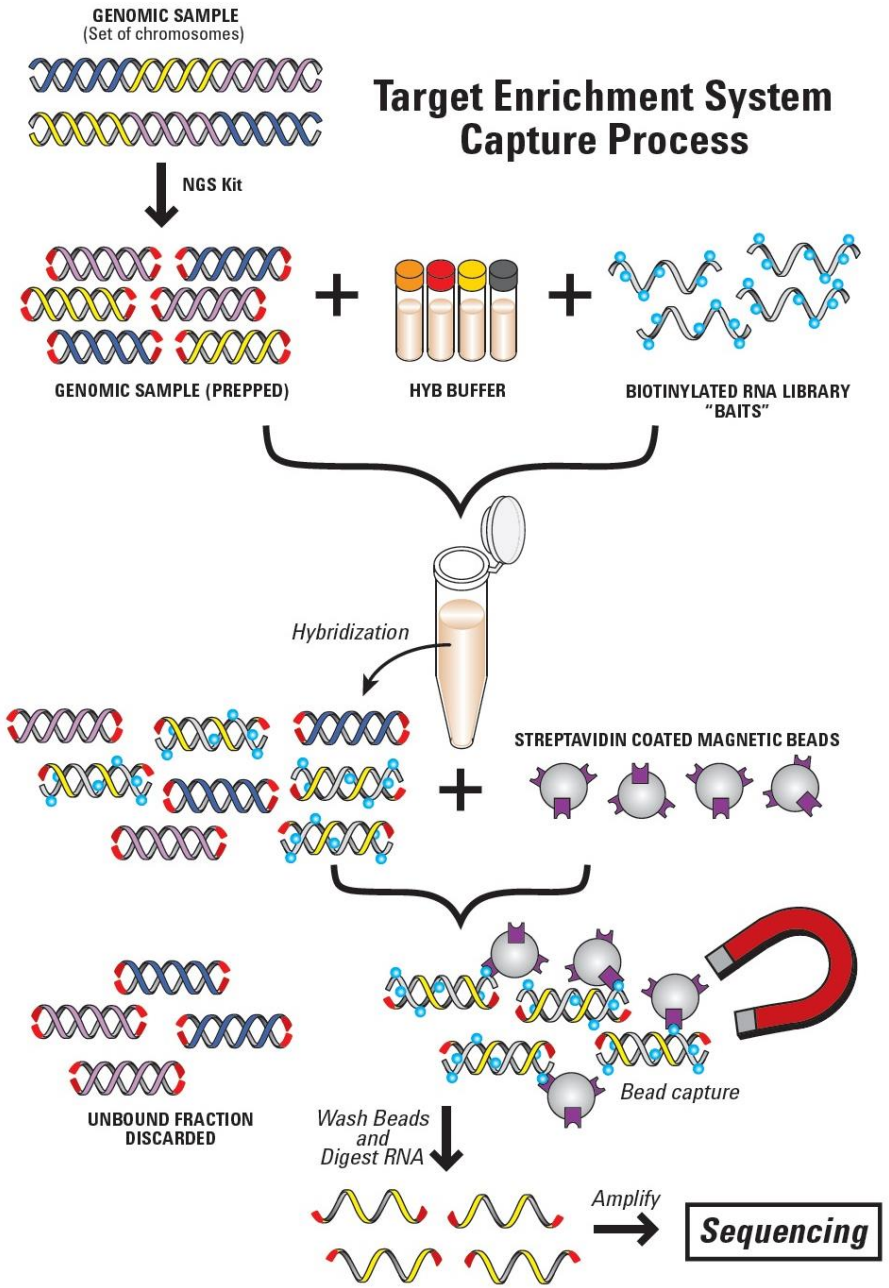


Image from Agilent.com

Where did the idea for sequence capture come from?

Western Blots

To allow detection of the target protein, the secondary antibody is commonly linked to **biotin**

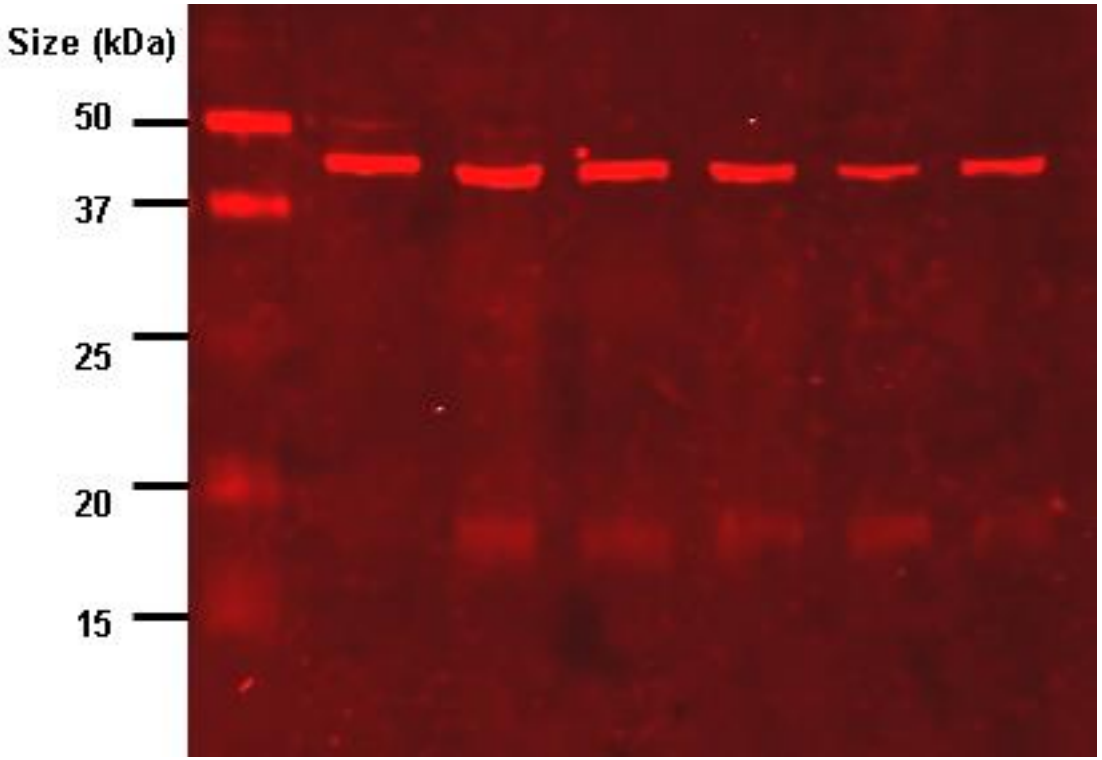
ELSEVIER

Analytical Biochemistry

Volume 152, Issue 2, 1 February 1986, Pages 329-332

Biotinylated proteins as molecular weight standards on Western blots

Dean Della-Penna, Rolf E. Christoffersen [1], Alan B. Bennett



Western Blots

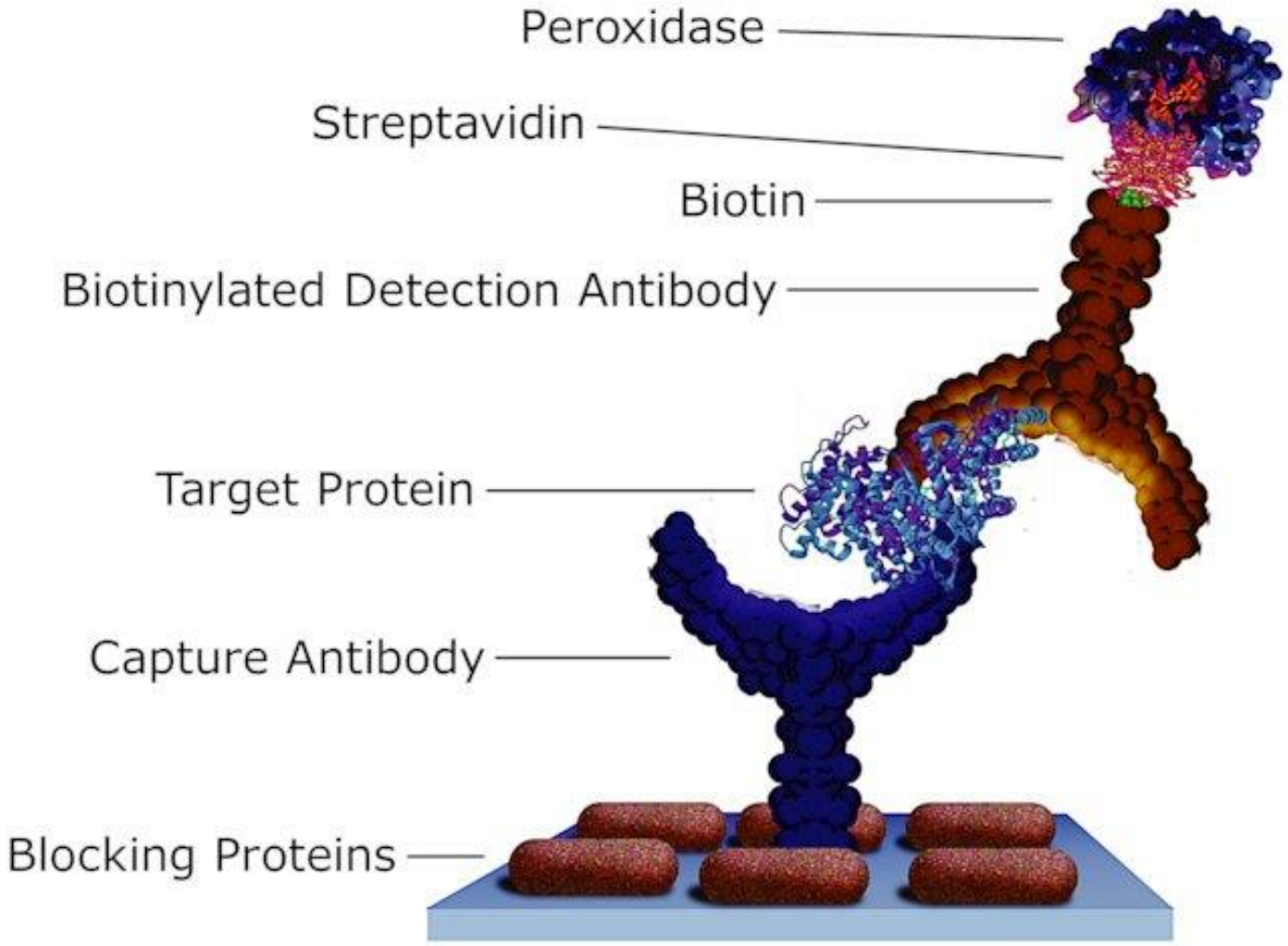


Image by Sigma Aldrich

Sequence Capture

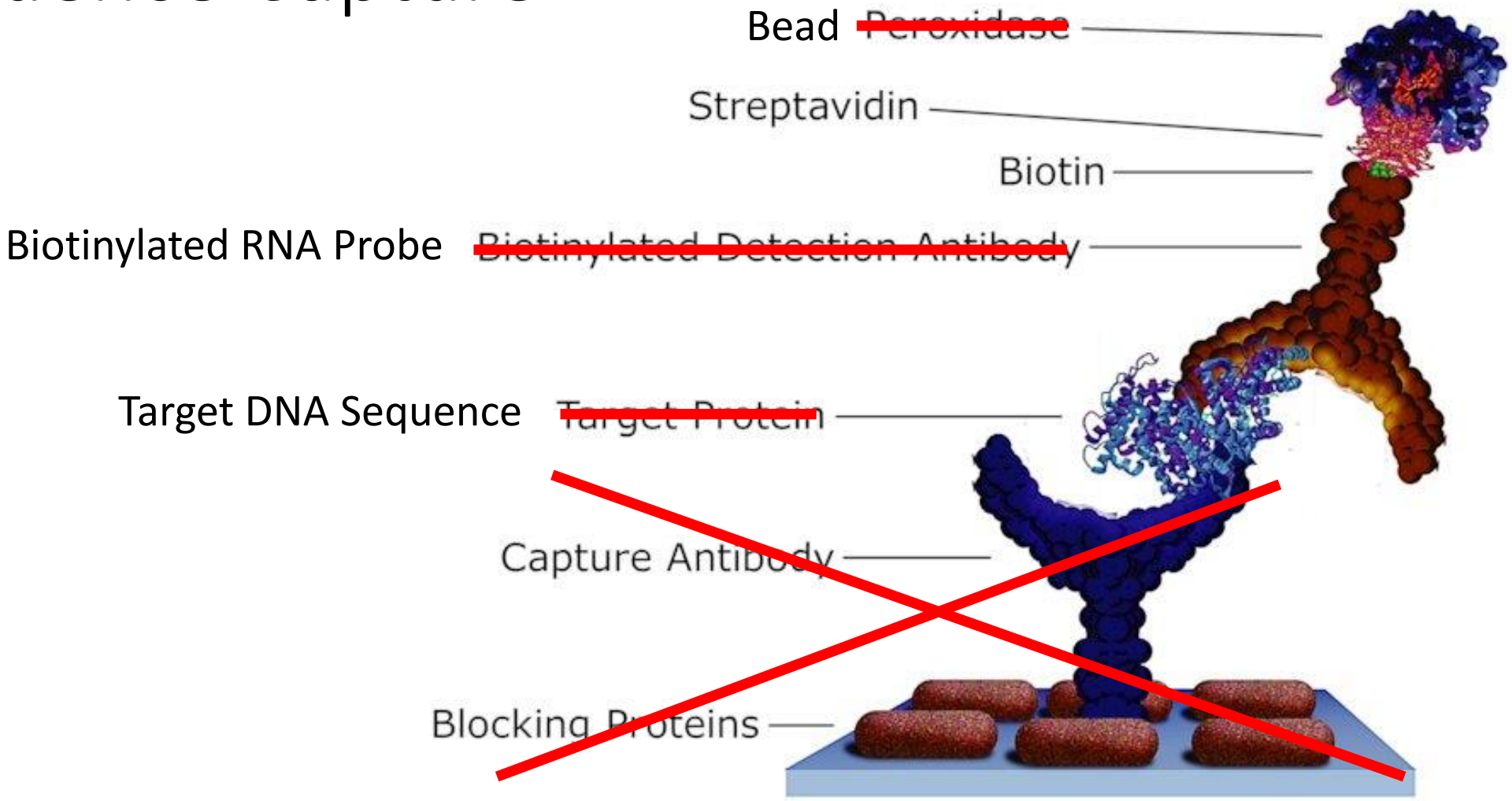


Image by Sigma Aldrich

Sequence Capture

Magnet

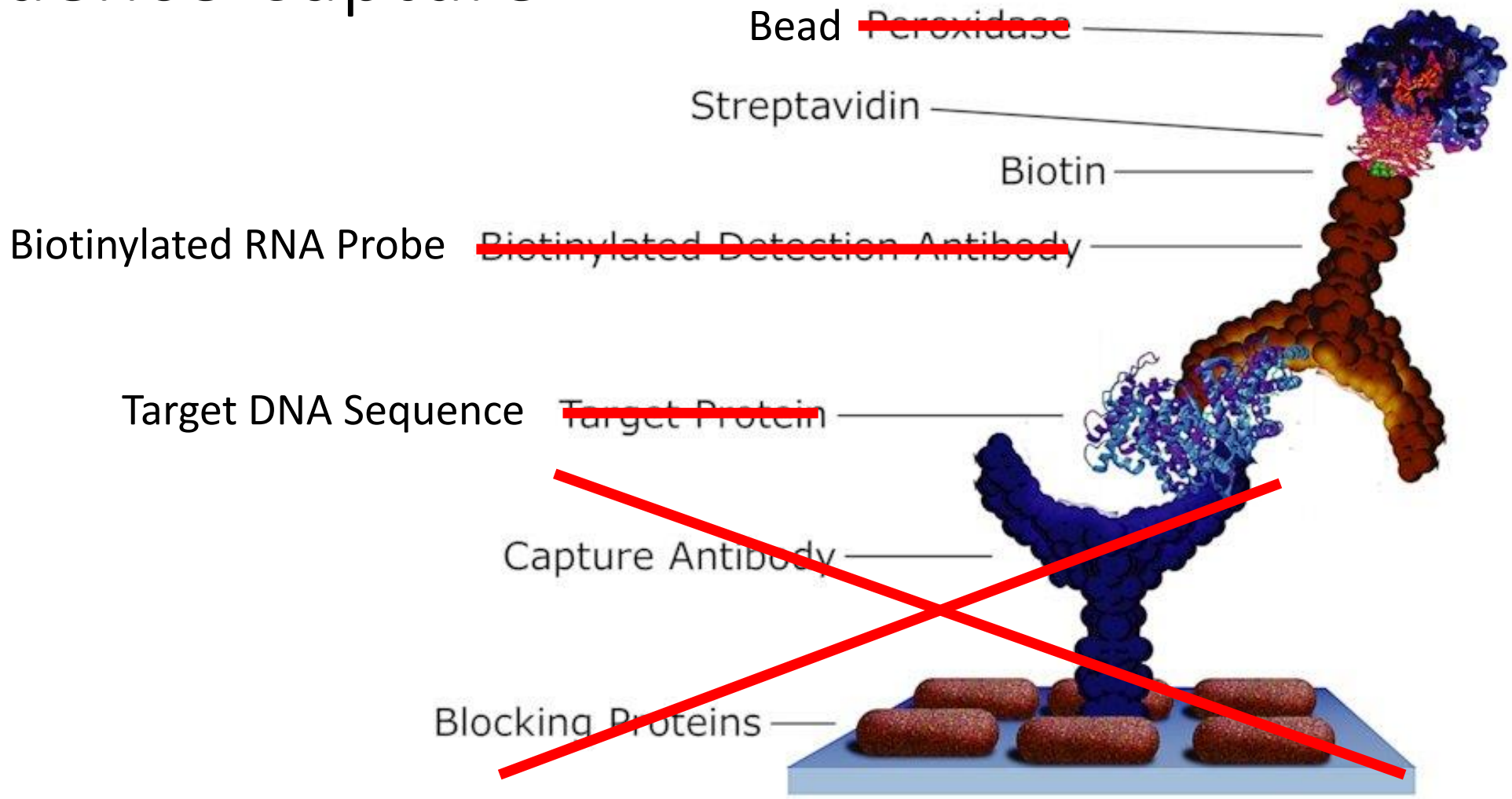


Image by Sigma Aldrich

Ultraconserved Elements (UCEs)

- First described in the human genome
- They are used to study…
- Genomics
- Phylogenetics

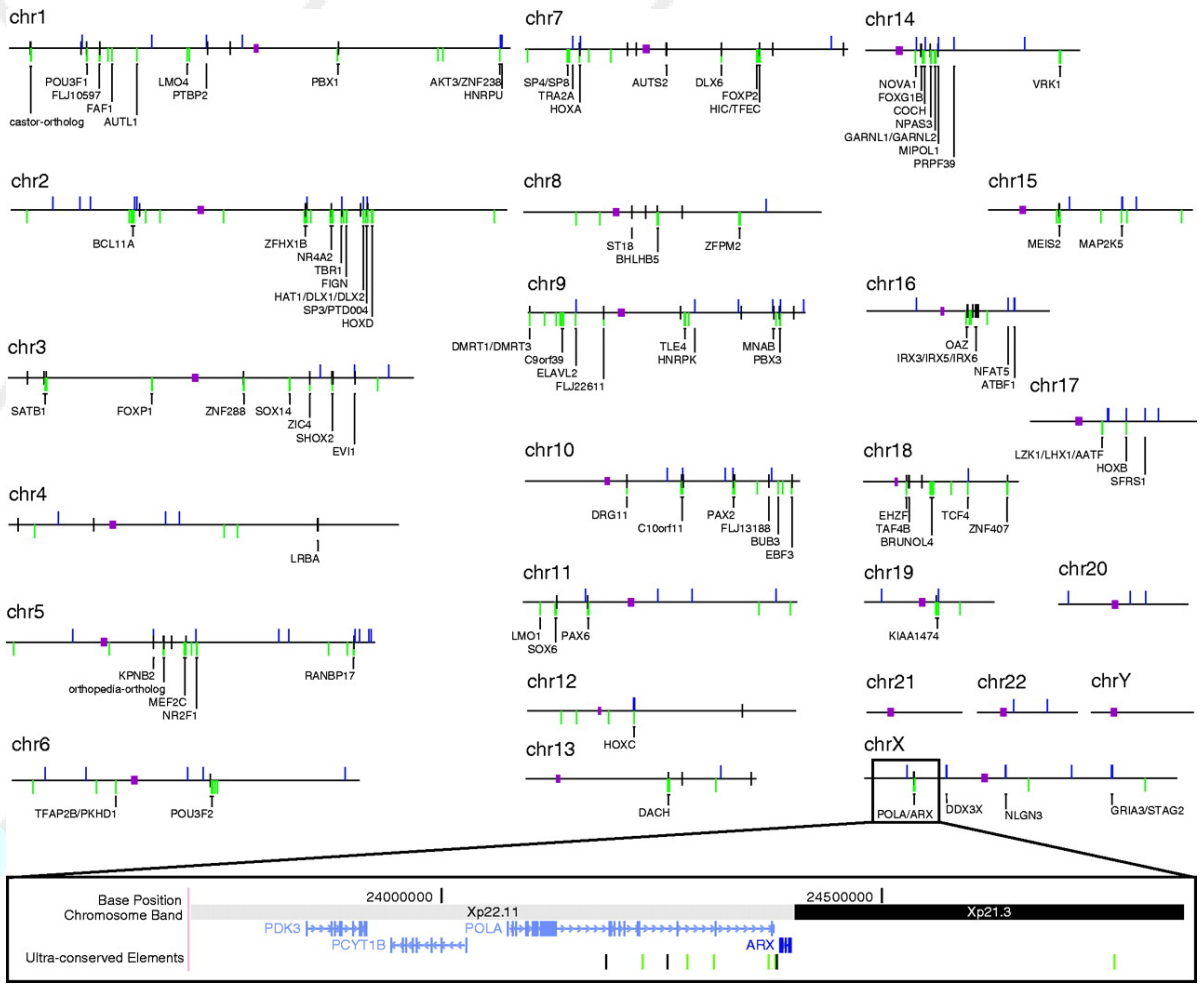
Text from Roche.com

What are UCEs?

Ultraconserved Elements in the Human Genome

Gill Bejerano,[1*] Michael Pheasant,[3] Igor Makunin,[3] Stuart Stephen,[3] W. James Kent,[1] John S. Mattick,[3] David Haussler[2*]

These ultraconserved elements of the human genome are most often located either overlapping exons in genes involved in RNA processing or in introns or nearby genes involved in the regulation of transcription and development



What are UCEs?

Resource

28-Way vertebrate alignment and conservation track in the UCSC Genome Browser

Webb Miller,[1,11] Kate Rosenbloom,[2] Ross C. Hardison,[1] Minmei Hou,[1] James Taylor,[3] Brian Raney,[2] Richard Burhans,[1] David C. King,[1] Robert Baertsch,[2] Daniel Blankenberg,[1] Sergei L. Kosakovsky Pond,[4] Anton Nekrutenko,[1] Belinda Giardine,[1] Robert S. Harris,[1] Svitlana Tyekucheva,[1] Mark Diekhans,[2] Thomas H. Pringle,[5] William J. Murphy,[6] Arthur Lesk,[1] George M. Weinstock,[7] Kerstin Lindblad-Toh,[8] Richard A. Gibbs,[7] Eric S. Lander,[8] Adam Siepel,[9] David Haussler,[2,10] and W. James Kent[2]

[1]Center for Comparative Genomics and Bioinformatics, Penn State University, University Park, Pennsylvania 16802, USA; [2]Center for Biomolecular Science and Engineering, University of California, Santa Cruz, California 95064, USA; [3]Courant Institute, New York University, New York, New York 10012, USA; [4]Antiviral Research Center, University of California at San Diego, San Diego, California 92103, USA; [5]Sperling Foundation, Eugene, Oregon 97405, USA; [6]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas 77843, USA; [7]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA; [8]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA; [9]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14853, USA; [10]Howard Hughes Medical Institute, Santa Cruz, California 95060, USA

Also identified more broadly in alignments of vertebrate genomes in 2007…

Why are there UCEs?

Genome organization and stability

Cell Reports
Article

OPEN ACCESS
CellPress

Ultraconserved Elements Occupy Specific Arenas of Three-Dimensional Mammalian Genome Organization

Ruth B. McCole,[1,2] Jelena Erceg,[1,2] Wren Saylor,[1] and Chao-ting Wu[1,3,*]
[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[2]These authors contributed equally
[3]Lead Contact
*Correspondence: twu@genetics.med.harvard.edu
https://doi.org/10.1016/j.celrep.2018.06.031

Why are there UCEs?

Conserve gene function as enhancers

BMC Genomics

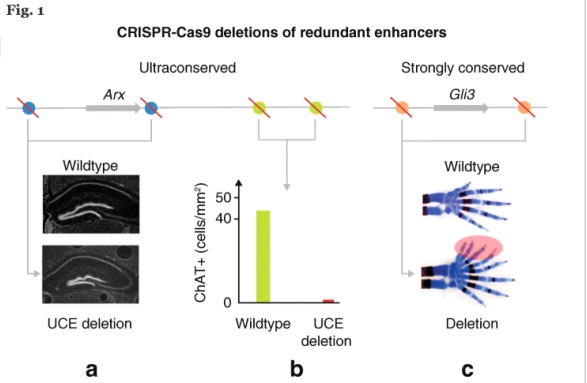
BioMed Central

Research article

Open Access

Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes

Albin Sandelin[†1], Peter Bailey[†2], Sara Bruce[1,3], Pär G Engström[1], Joanna M Klos[2], Wyeth W Wasserman[4], Johan Ericson[*2] and Boris Lenhard[*1]



Pairwise deletion of redundant ultraconserved elements in the locus of the mouse *Arx* gene [4] (**a**, **b**) and redundant strongly conserved enhancers of the mouse *Gli3* gene [8] (**c**). A combined deletion of two dorsal forebrain enhancers hs122 and hs123 leads to a smaller dentate gyrus (*white staining*) with disorganized appearance (**a**). A combined deletion of two ventral forebrain enhancers hs119 and hs121 leads to a drastic decrease in the density of striatal cholinergic neuron density (**b**). A combined deletion of two *Gli3* limb enhancers in a sensitized genetic background leads to a severe polydactyly (**c**). *ChAT* choline acetyltransferase, *UCE* ultraconserved element

Elnitski & Ovcharenko, 2018
Genome Biology

Controversy?

News & Views | Published: 29 March 2021

FUNCTIONAL GENOMICS

Ultraconservation of enhancers is not ultranecessary

Maureen Pittman & Katherine S. Pollard

Nature Genetics **53**, 429–430 (2021) | Cite this article

2405 Accesses | **20** Altmetric | Metrics

Research Highlight | Open Access | Published: 08 May 2018

The hypothesis of ultraconserved enhancer dispensability overturned

Laura Elnitski & Ivan Ovcharenko

Genome Biology **19**, Article number: 57 (2018) | Cite this article

3089 Accesses | **5** Altmetric | Metrics

Ultraconserved Elements (UCEs)

- UCEs as phylogenetic markers
- In vertebrate systems this largely started with Faircloth et al. (2012)

Phylogenetics

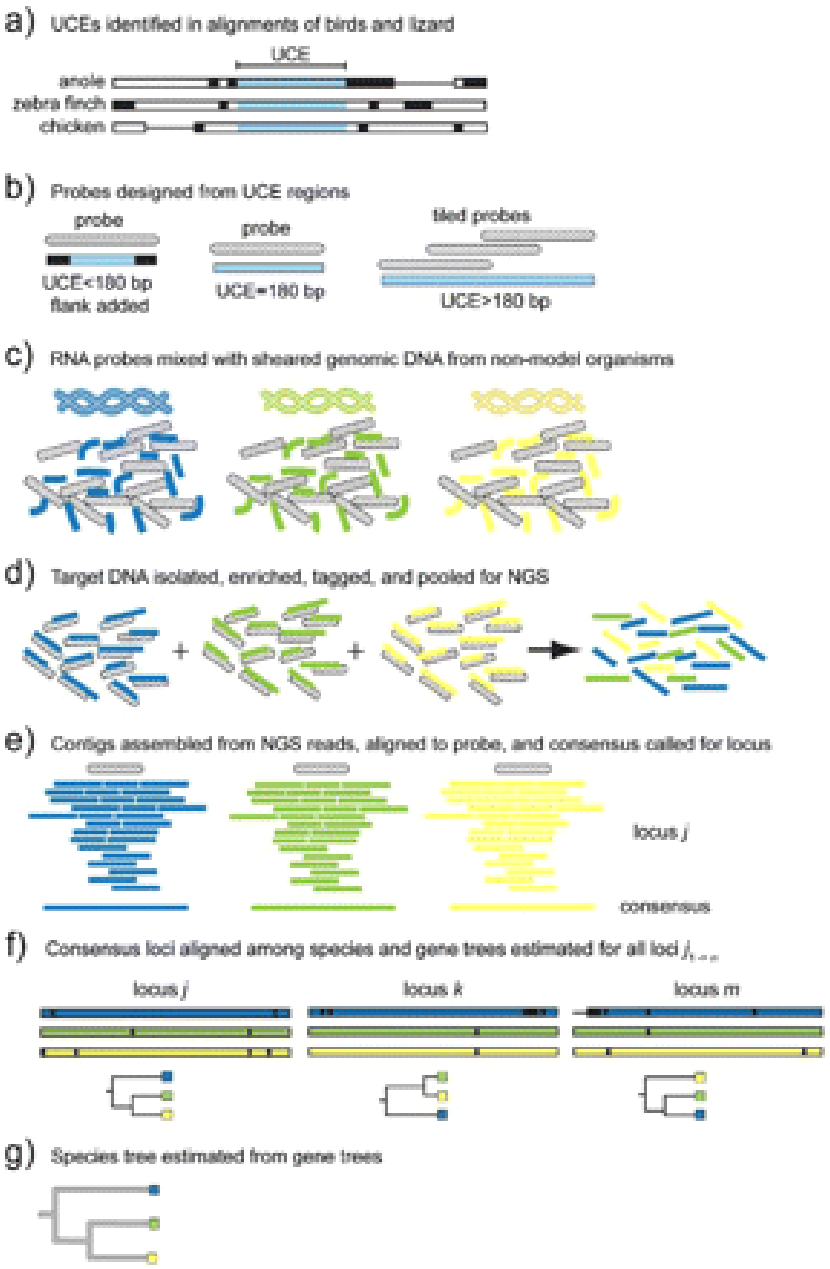
Syst. Biol. 61(5):717–726, 2012

DOI:10.1093/sysbio/sys004
Advance Access publication on January 9, 2012

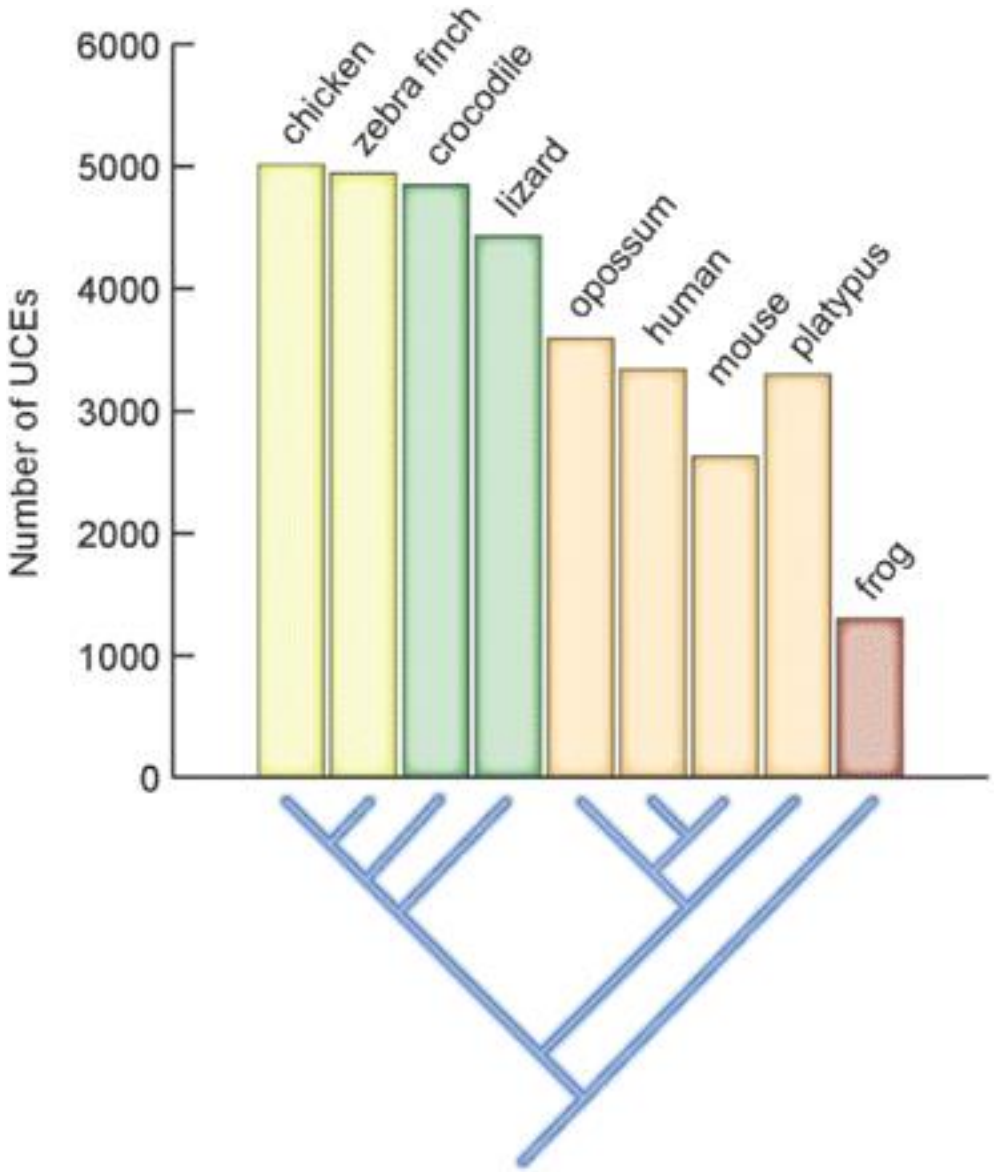
Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales

BRANT C. FAIRCLOTH[1,*], JOHN E. MCCORMACK[2], NICHOLAS G. CRAWFORD[3], MICHAEL G. HARVEY[2,4], ROBB T. BRUMFIELD[2,4], AND TRAVIS C. GLENN[5]

[1]*Department of Ecology and Evolutionary Biology, 621 Charles E. Young Drive, University of California, Los Angeles, CA 90095, USA;* [2]*Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA;* [3]*Department of Biology, Boston University, Boston, MA 02215, USA;* [4]*Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA; and* [5]*Department of Environmental Health Science and Georgia Genomics Facility, University of Georgia, Athens, GA 30602, USA;*
[*]*Correspondence to be sent to: Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA; E-mail: brant@faircloth-lab.org.*

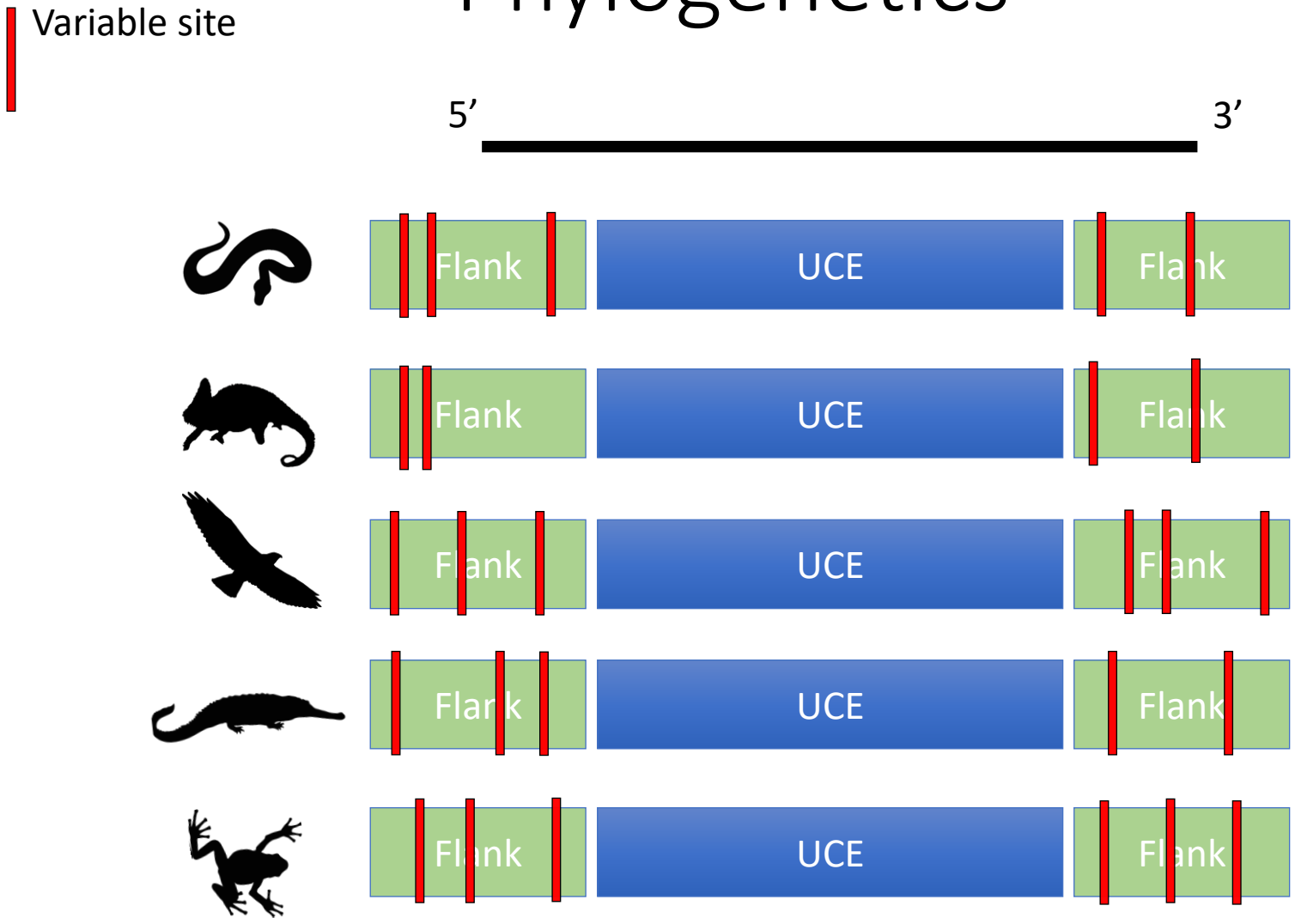


From Faircloth et al. 2012, Syst. Biol.

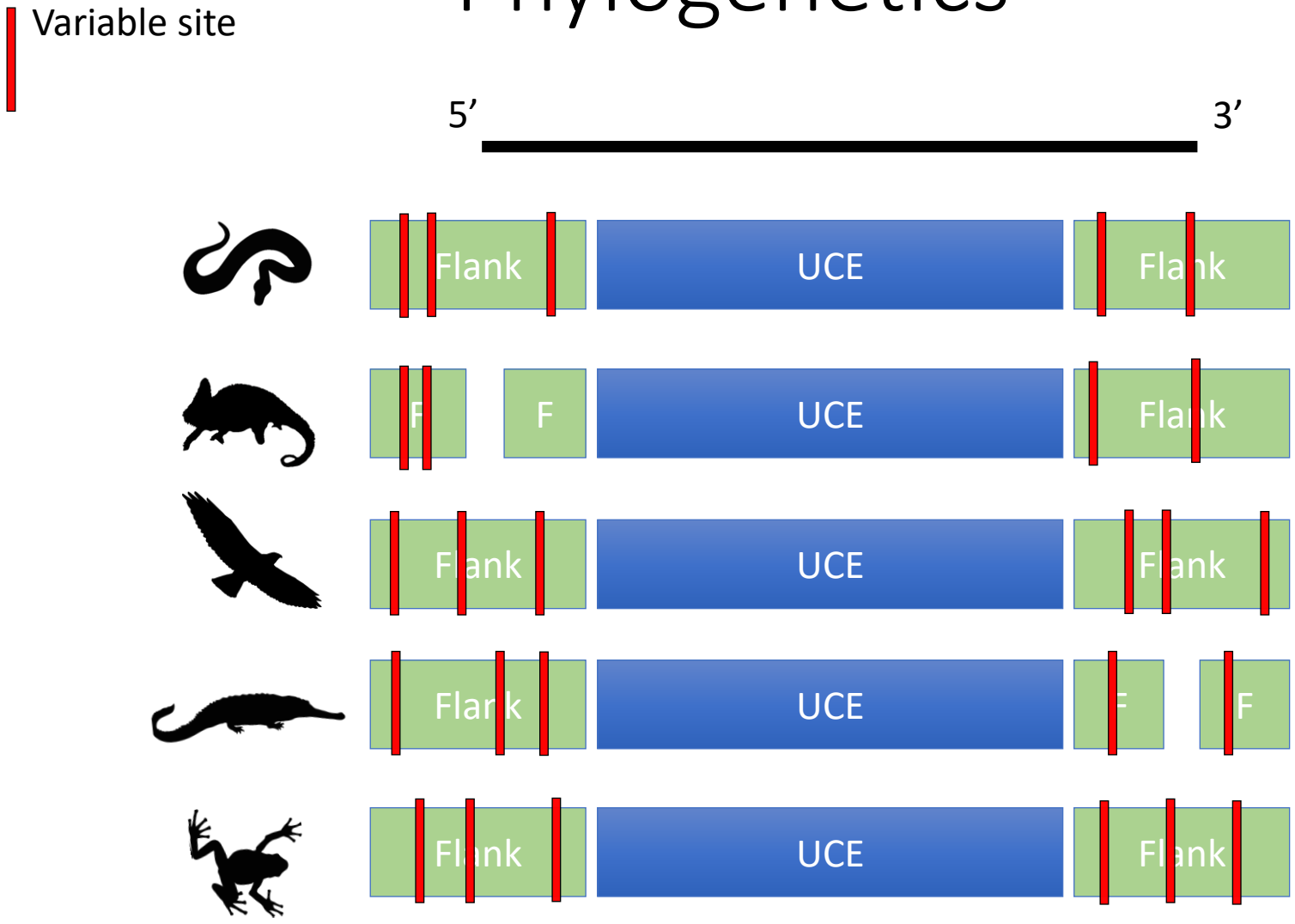


From Faircloth et al. 2012, Syst. Biol.

Phylogenetics



Phylogenetics



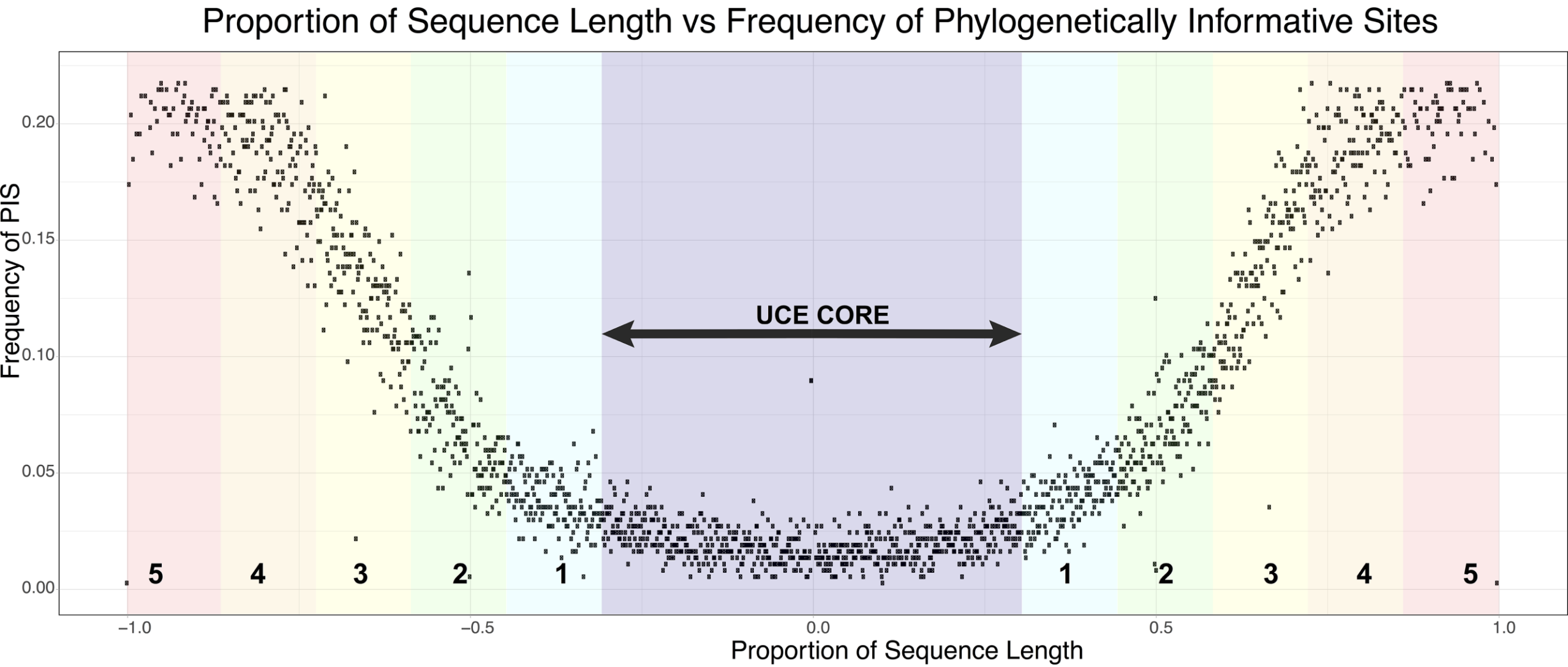


Image from Van Dam et al. 2017, PLoS ONE

Phylogenomic species delimitation and host-symbiont coevolution in the fungus-farming ant genus *Sericomyrmex* Mayr (Hymenoptera: Formicidae): ultraconserved elements (UCEs) resolve a recent radiation

ANA JEŠOVNIK[1,2], JEFFREY SOSA-CALVO[1,3,4], MICHAEL W. LLOYD[1], MICHAEL G. BRANSTETTER[1,5], FERNANDO FERNÁNDEZ[6] and TED R. SCHULTZ[1,2]

[1]Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC, U.S.A., [2]Maryland Center for Systematic Entomology, Department of Entomology, University of Maryland, College Park, MD, U.S.A., [3]University of Rochester, Rochester, NY, U.S.A., [4]School of Life Sciences, Arizona State University, AZ, U.S.A., [5]University of Utah, Salt Lake City, UT, U.S.A. and [6]Universidad Nacional de Colombia, Bogotá D.C., Colombia

biology letters
Evolutionary biology
Biol. Lett. (2012) **8**, 783–786
doi:10.1098/rsbl.2012.0331
Published online 16 May 2012

More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs

Nicholas G. Crawford[1,*], Brant C. Faircloth[2], John E. McCormack[3], Robb T. Brumfield[3,4], Kevin Winker[5] and Travis C. Glenn[6]

[1]Department of Biology, Boston University, Boston, MA 02215, USA
[2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA
[3]Museum of Natural Science, and [4]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA
[5]University of Alaska Museum, 907 Yukon Drive, Fairbanks, AK 99775, USA
[6]Department of Environmental Health Science and Georgia Genomics Facility, University of Georgia, Athens, GA 30602, USA
**Author for correspondence (ngcrawford@gmail.com).*

Method

Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis

John E. McCormack,[1,8] Brant C. Faircloth,[2] Nicholas G. Crawford,[3] Patricia Adair Gowaty,[4,5] Robb T. Brumfield,[1,6] and Travis C. Glenn[7]

[1]Museum of Natural Science, Louisiana State University, Baton Rouge, Louisiana 70803, USA; [2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 90095, USA; [3]Department of Biology, Boston University, Boston, Massachusetts 02215, USA; [4]Smithsonian Tropical Research Institute, Apartado Postal 0843-03092, Panamá, República de Panamá; [5]Institute of the Environment, University of California, Los Angeles, California 90095, USA; [6]Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA; [7]Department of Environmental Health Science, University of Georgia, Athens, Georgia 30602, USA

Blaimer *et al. BMC Evolutionary Biology* (2015) 15:271
DOI 10.1186/s12862-015-0552-5
BMC Evolutionary Biology

RESEARCH ARTICLE — Open Access

Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants

Bonnie B. Blaimer[1*], Seán G. Brady[1], Ted R. Schultz[1], Michael W. Lloyd[1], Brian L. Fisher[2] and Philip S. Ward[3]

PLOS | ONE

RESEARCH ARTICLE

Ultraconserved elements (UCEs) resolve the phylogeny of Australasian smurf-weevils

Matthew H. Van Dam[1¤*], Athena W. Lam[1¤], Katayo Sagata[2], Bradley Gewa[3], Raymond Laufa[3], Michael Balke[1,4], Brant C. Faircloth[5], Alexander Riedel[6]

1 SNSB-Zoological State Collection, Münchhausenstraße 21, München, Germany, **2** School of Natural & Physical Sciences, The University of Papua New Guinea, UNIVERSITY 134, National Capital District, Papua New Guinea, **3** The New Guinea Binatang Research Center, Madang, Papua New Guinea, **4** GeoBioCenter, Ludwig-Maximilians-Universität, München, Germany, **5** Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA, United States of America, **6** State Museum of Natural History Karlsruhe, Karlsruhe, Germany

¤ Current address: Entomology Department, California Academy of Sciences, San Francisco, CA, United States of America.
* matthewhvandam@gmail.com

Evolutionary biology

Phylogenomic analyses of more than 4000 nuclear loci resolve the origin of snakes among lizard families

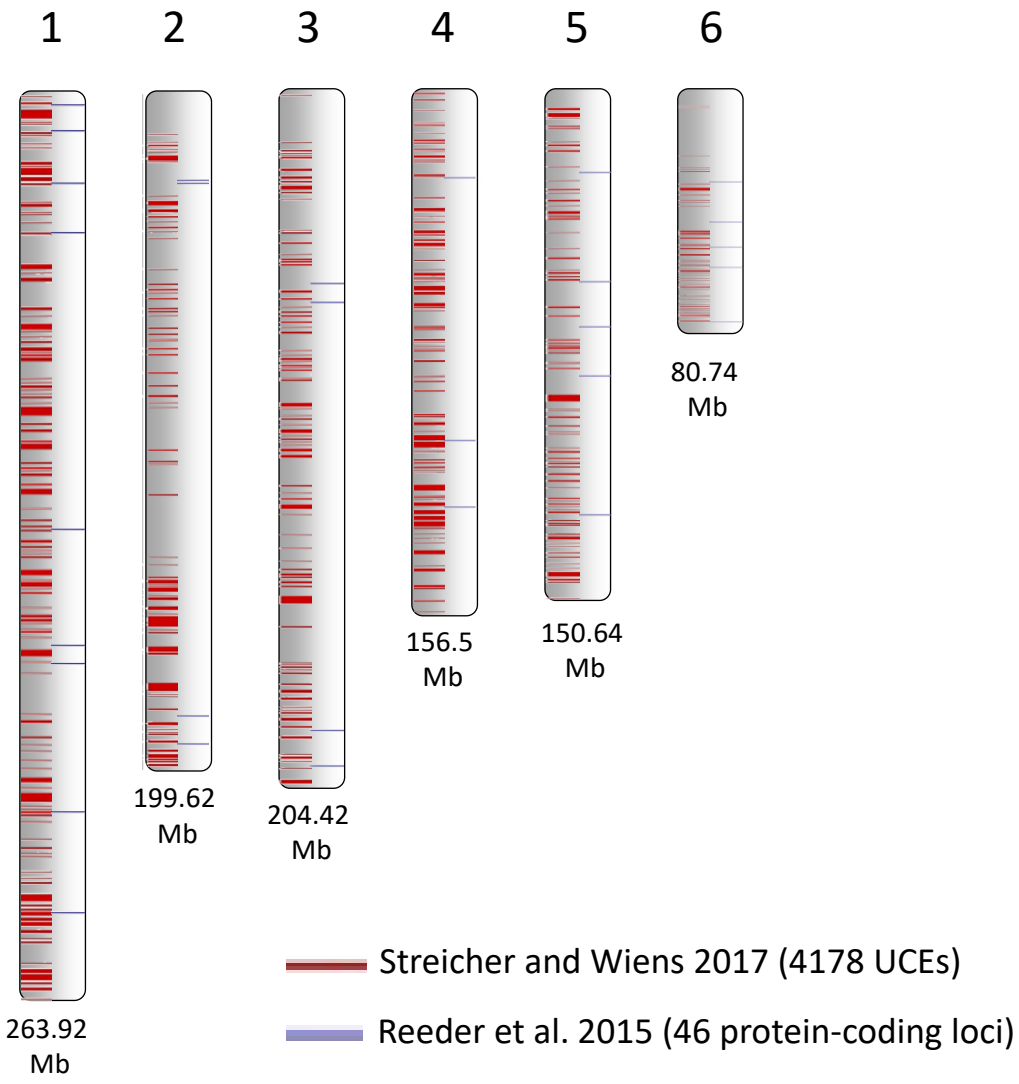
Jeffrey W. Streicher[1,2] and John J. Wiens[1]

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721-0088, USA
[2]Department of Life Sciences, The Natural History Museum, London SW7 5BD, UK

Empirical Studies: Qualities of UCEs

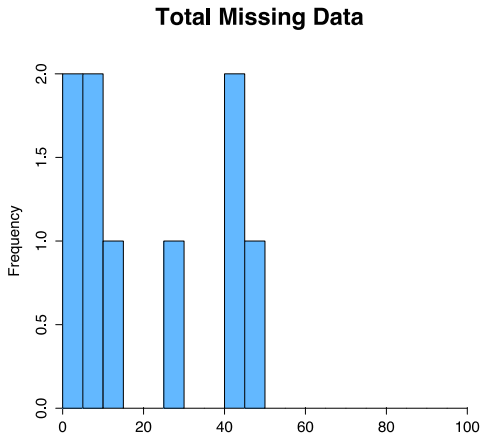
- Iguanian Lizards
- > 3,000 UCEs
- 44 taxa

Anolis carolinensis
Macro-chromosomes

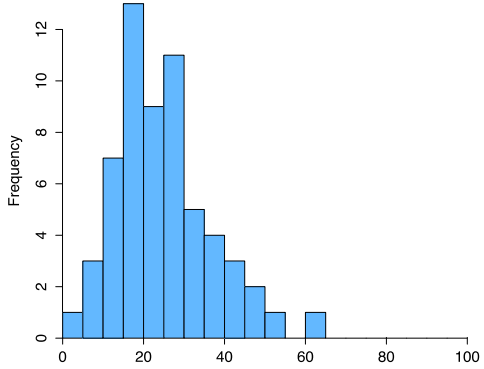


GENOME DISTRIBUTION

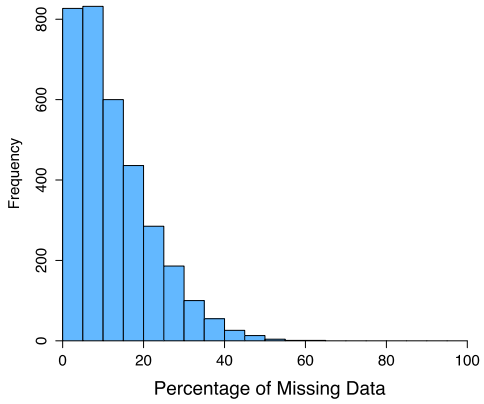
mtDNA



Protein-coding nucDNA



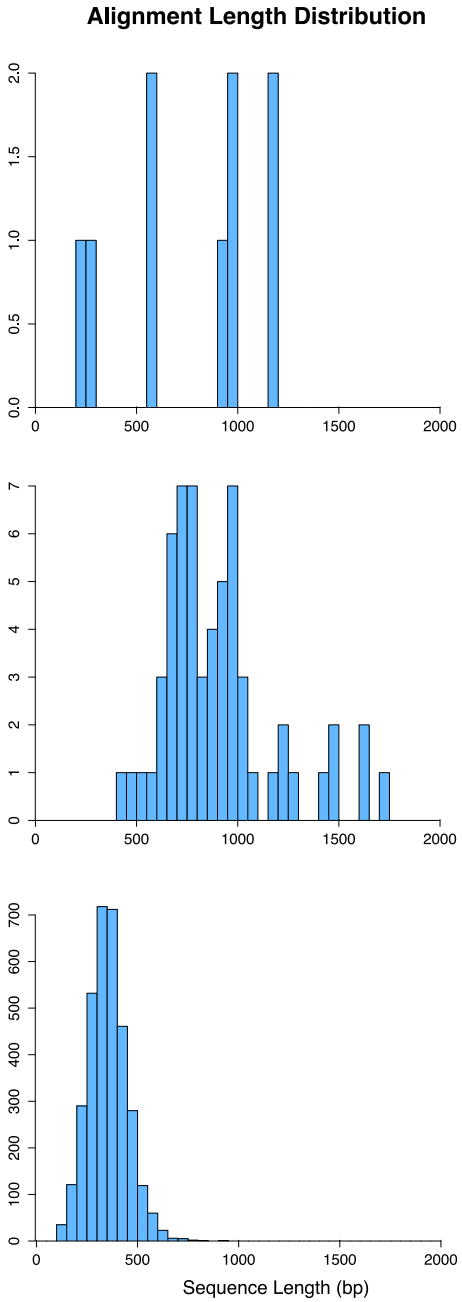
UCEs nucDNA



MISSING DATA

Portik et al. In Prep.

mtDNA



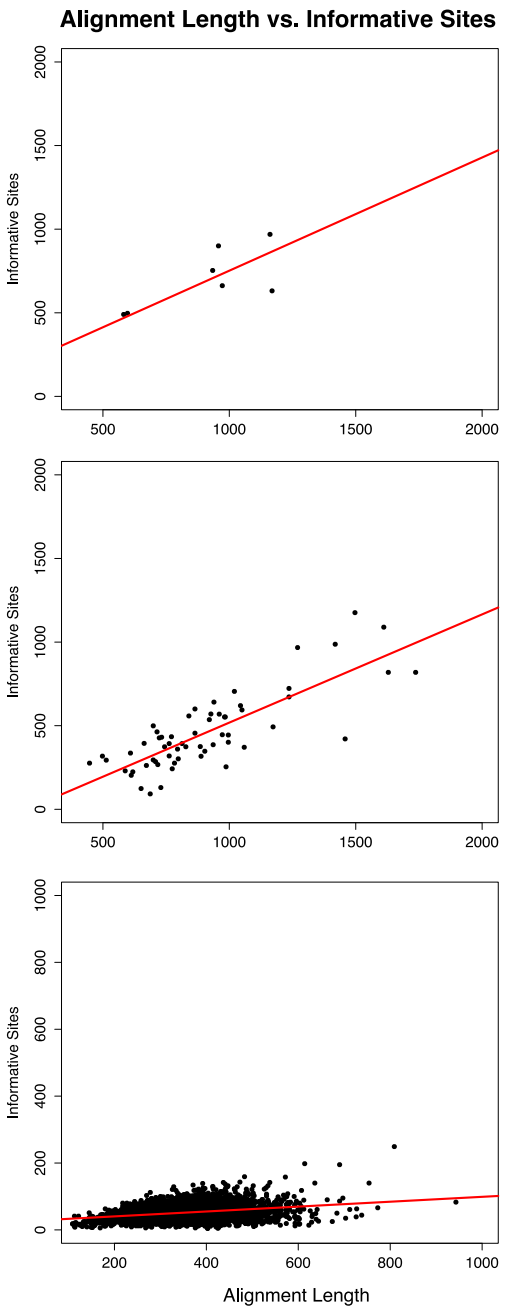
Protein-coding nucDNA

UCEs nucDNA

LENGTH

Portik et al. In Prep.

mtDNA



Protein-coding nucDNA

UCEs nucDNA

VARIABILITY

Portik et al. In Prep.

Ultraconserved Elements (UCEs)

- Useful for comparing/aligning genomes of different species
- Useful for phylogenetic analysis across divergent species
- Not as useful for population genetic analysis… but stay tuned for Unit 4 - ddRADseq

Unit 3: Targeted sequence capture of ultraconserved elements (UCEs)

Bioinformatics Lab

https://github.com/nhm-herpetology/museum-NGS-training

Getting ready…

- cd NGS_course
- Mkdir Unit_3
- cd Unit_1
- cd sratoolkit.2.11.1-ubuntu64
- cd bin