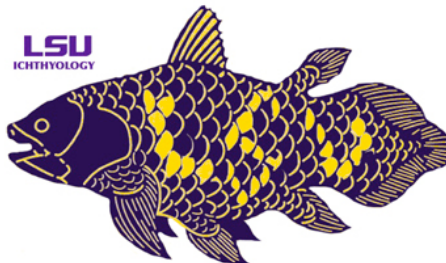


Phylogenetic insights into the evolution of Ostariophysi using Ultraconserved Elements



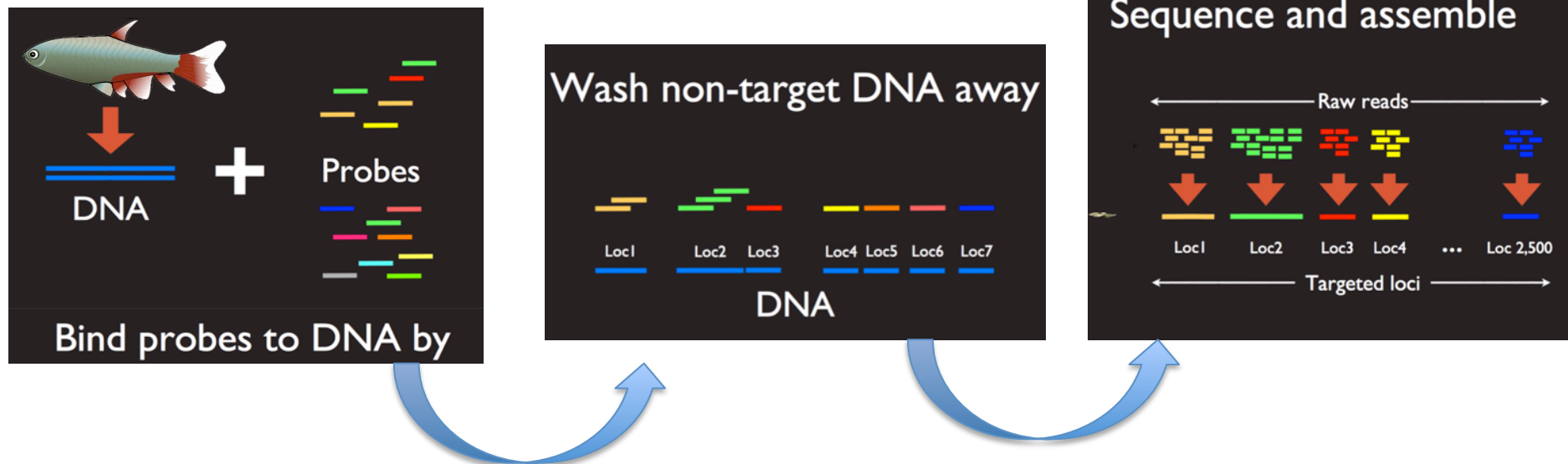
Prosanta Chakrabarty, Caleb McMahan,
Bill Fink, James Albert, Jairo Arroyave,
Melanie Stiassny, Brant Faircloth, Mike Alfaro



Sequence Capture



- Next-gen sequencing much more effective than traditional Sanger sequencing
 - rapidly obtain hundreds of loci, for relatively low cost
- Genome reduction (Genomes are billions of bp)



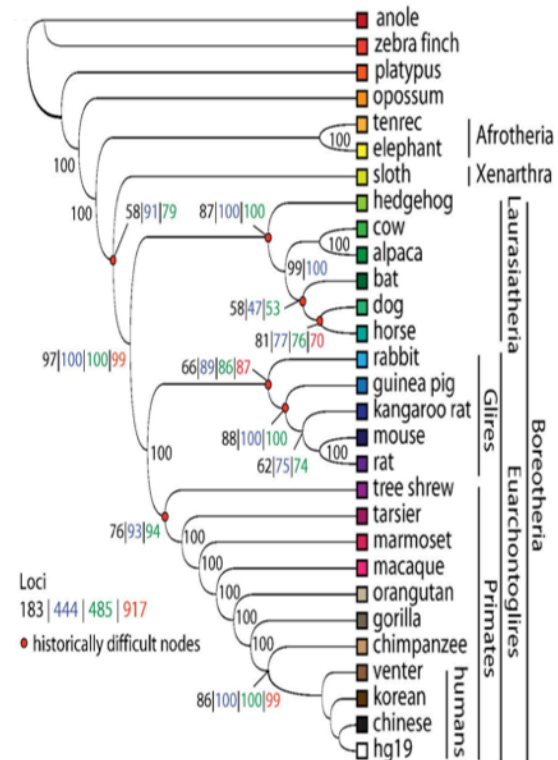
Massively Parallel Sequencing of Ultraconserved Elements (UCEs)

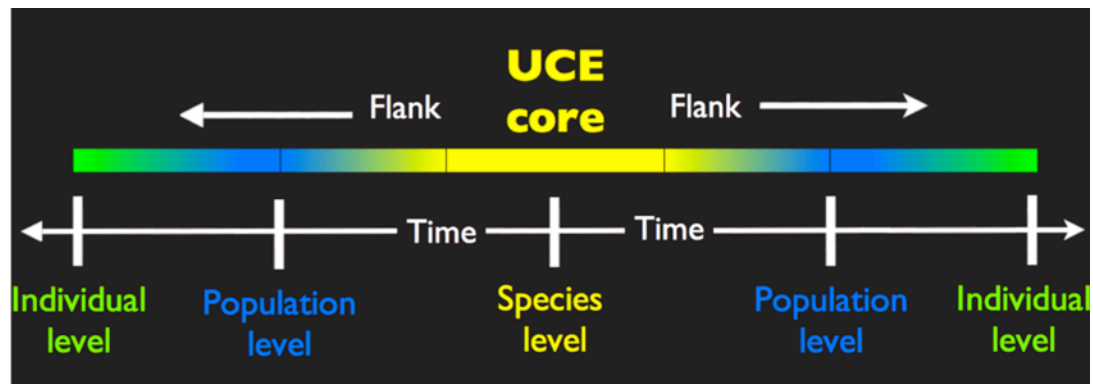
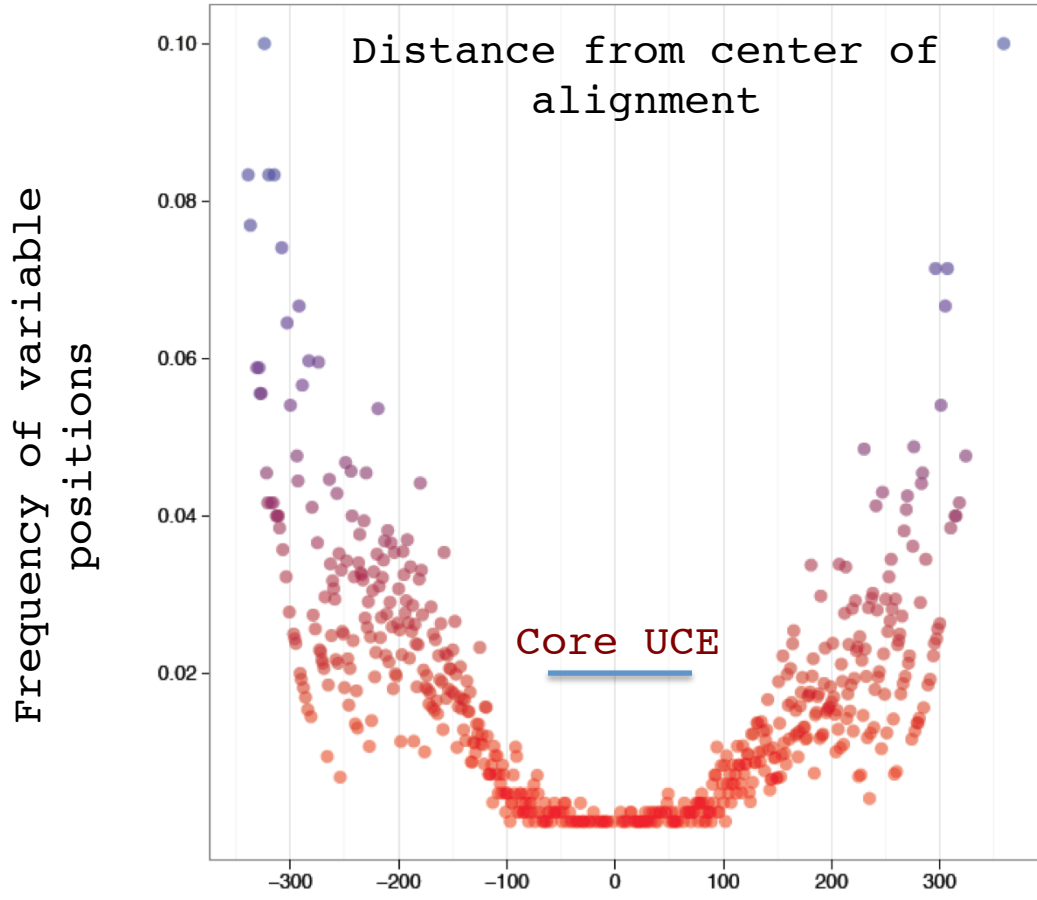
- UCEs are universal homologous genomic markers, present across vertebrates (90-95% conserved)

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*}

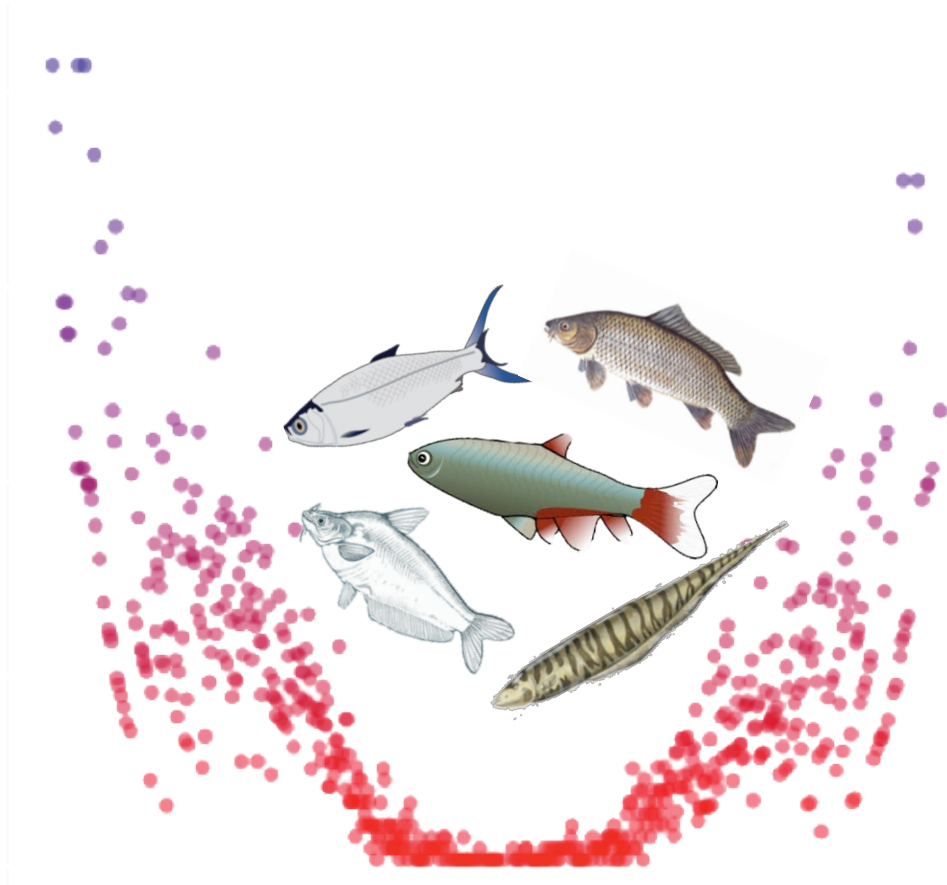
There are 481 segments longer than 200 base pairs (bp) that are absolutely conserved (100% identity with no insertions or deletions) between orthologous regions of the human, rat, and mouse genomes. Nearly all of these segments are also conserved in the chicken and dog genomes, with an average of 95 and 99% identity, respectively. Many are also significantly conserved in fish. 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. Along with more than 5000 sequences of over 100 bp that are absolutely conserved among the three sequenced mammals, these represent a class of genetic elements whose functions and evolutionary origins are yet to be determined, but which are more highly conserved between these species than are proteins and appear to be essential for the ontology of mammals and other vertebrates.





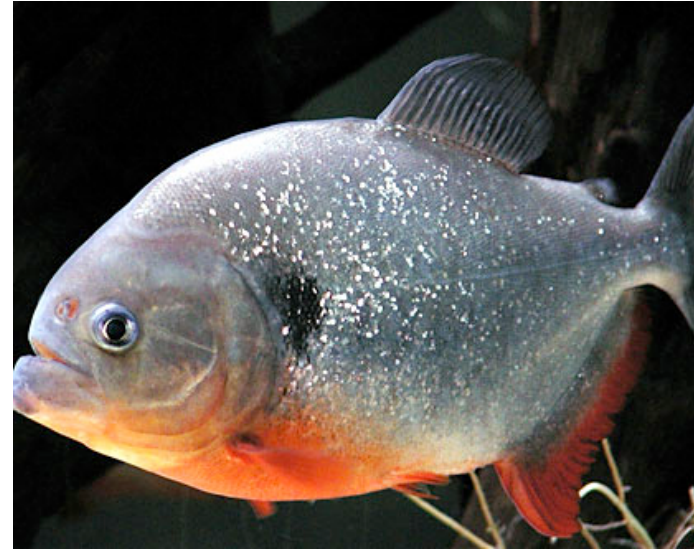
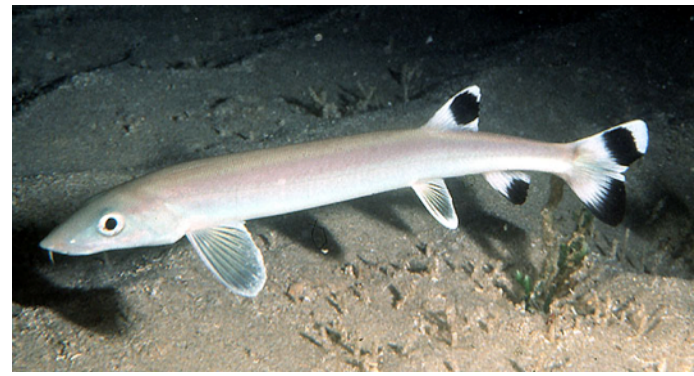
One of the Toughest Questions in Ichthyology

What are the basal relationships of ostariophysans?



Ostariophysi

- Anotophysi
 - Gonorhynchiformes (37 spp.)
- Otophysi
 - Cypriniformes (4000 spp.)
 - Characiformes (2000 spp.)
 - Siluriformes (3500 spp.)
 - Gymnotiformes (200 spp.)



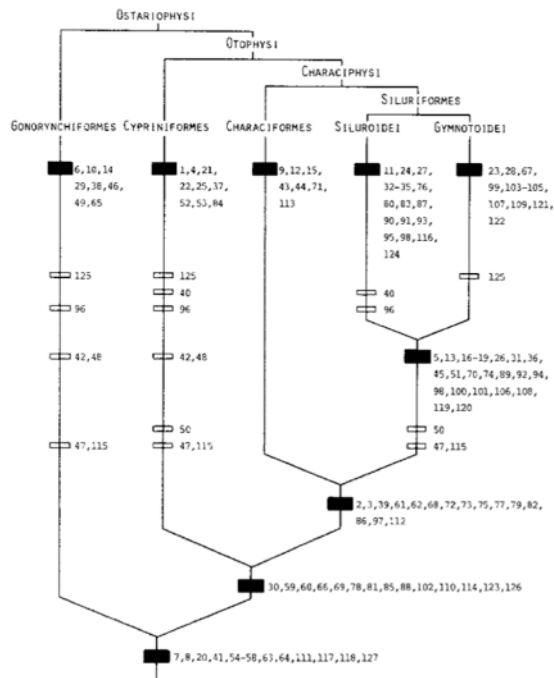
Ostariophysii

- 10,000 spp. (30% of all teleosts, 20% of all vertebrates)
- 70% of all freshwater fishes

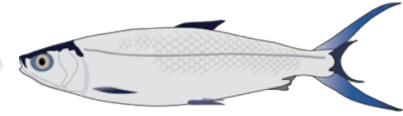


- Include milkfish, minnows, carps, barbs, suckers, loaches, piranhas, tetras, catfishes, knifefishes, electric eels
- United by Weberian apparatus, schreckstoff, among other characters

Fink & Fink 1981



GONORHYNCHIFORMES



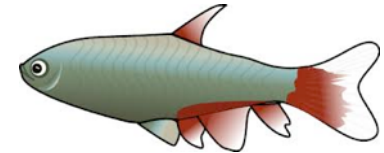
Ostariophysi

CYPRINIFORMES



Otophysi

CHARACIFORMES



Characiphysi

GYMNOTIFORMES



Siluriphysi

SILURIFORMES

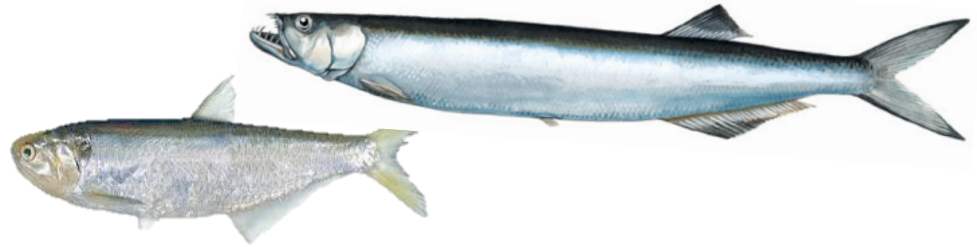


Morphology, 127 characters (only 8 homoplasious)

- **Distant Outgroups** – *Lepisosteus oculatus* [root], *Gadus morhua*, *Oryzias latipes*, *Umbra limi*, *Diaphus theta*

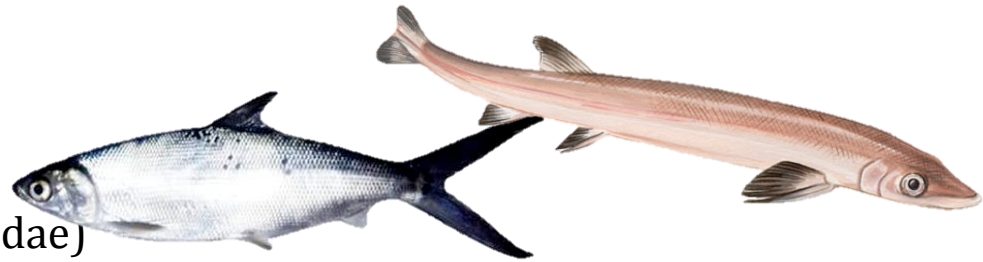
- **Outgroup [Otocephala]- Clupeiformes**

- *Thryssa hamiltoni* (Engraulidae)
- *Chirocentrus dorab* (Chirocentridae)
- *Dorosoma petenense* (Clupeidae)



- **Gonorhynchiformes**

- *Gonorhynchus* sp. (Gonorhynchidae)
- *Chanos chanos* (Chanidae)
- *Phractolaemus ansorgii* (Phractolaemidae)
- *Parakneria abbreviate* (Kneriidae)



- **Cypriniformes**

- *Moxostoma poecilurum* (Catosomidae)
- *Danio rerio*, *Cyprinella venusta*, *Puntioplites falcifer* (Cyprinidae)



- **Gymnotiformes**

- *Apteronotus albifrons* (Apteronotidae)
- *Sternopygus macrurus* (Sternopygidae)
- *Brachyhyopomus occidentalis* (Hypopomidae)
- *Gymnotus cylindricus*, *G. carapo* (Gymnotidae)



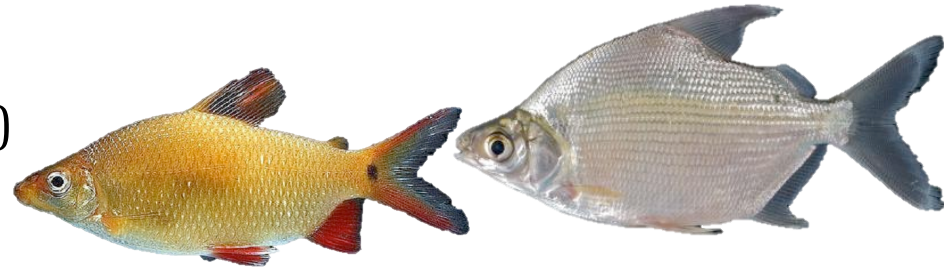
- **Characiformes Neotropics**

- *Astyanax aeneus* (Characidae)
- *Ctenolucius beani* (Ctenoluciidae)
- *Hoplias microlepis* (Erythrinidae)



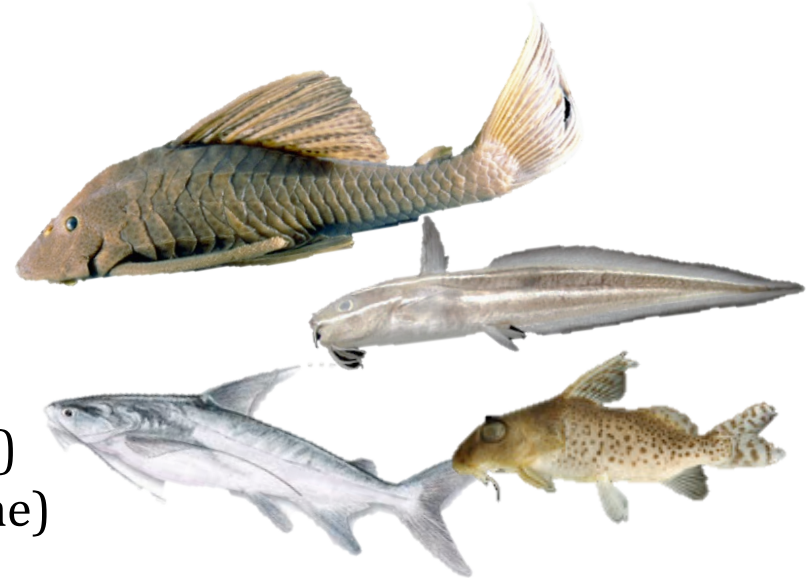
- **Characiformes Africa**

- *Citharinus gibbosus* (Citharinidae)
- *Brycinus macrolepidotis* (Alestidae)
- *Hepsetus odoe* (Hepsetidae)
- *Bryconaethiops yseuxi* (Alestidae)
- *Xenocharax crassus*, *Nannaethiops unitaeniatus*, *Distichodus hypostomatus*, *D. maculatus* (Distichodontidae)



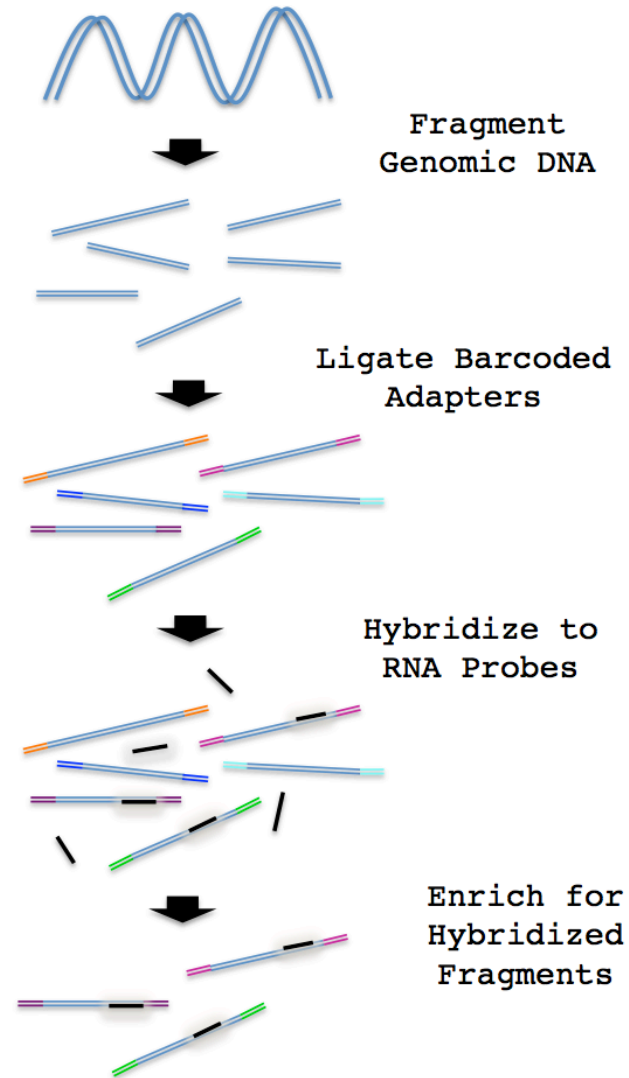
- **Siluriformes**

- *Hypostomus panamensis* (Loricariidae)
- *Plotosus lineatus* (Plotosidae)
- *Arius felis* (Ariidae)
- *Ameiurus natalis* (Ictaluridae)
- *Pangasius cf. pangasius* (Pangasiidae)
- *Auchenoglanis occidentalis* (Claroteidae)
- *Malapterurus stiasnyae* (Malapteruridae)
- *Synodontis filamentosus* (Mochokidae)



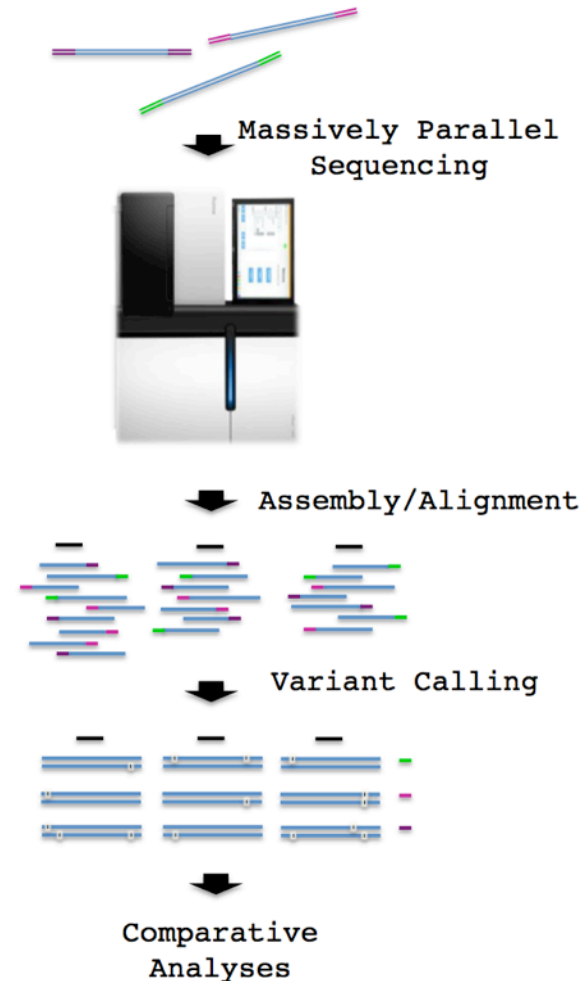
UCE Simplified Methods 1

- Qiagen DNEasy extraction
- Genomes sheared with sonicator
- **TruSeq-compatible library prep with custom indexes (add barcodes to sequences that tag individuals)**
- Pooled-sample (8 tagged libraries) target enrichment with probes targeting approximately 1,200 UCEs (**multiplex samples; add UCE primers “probes”**)
- **Post-enrichment PCR** amplification of 16-20 cycles
- **qPCR quantitation and normalization** (a check to see if amplification worked)



UCE Simplified Methods 2

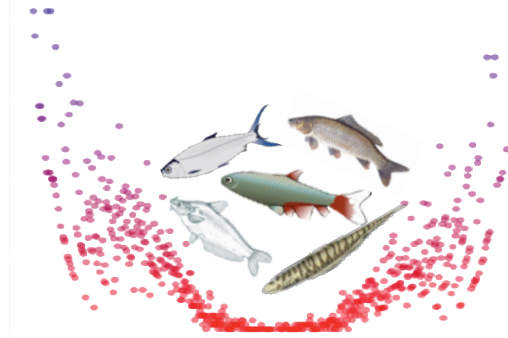
- **Next-gen Sequencing** - sequencing of enriched libraries on **HiSeq and MiSeq**
- Illumina Casava 1.8.2 to **demultiplex** (separate sequences by barcodes)
- read data adapter- and quality- trimmed <https://github.com/faircloth-lab/illumiprocessor> (**remove adaptors**)
- automated **assembly** using Trinity <http://trinityrnaseq.sourceforge.net/>
- identification and processing using <https://github.com/faircloth-lab/phyluce>
 - duplicate locus identification and removal
 - Map UCE's to particular part of genome
 - alignment using mafft
 - merged all individual nexus files from alignment
 - **Parallel RaxML analysis of data w/ GTRGAMMA; 100 bootstrap replicates**



protocols/code all open-source/open-access -
<http://ultraconserved.org>

Results

- Kept all UCEs that had at least 75% of taxa represented
- **351 loci** of mean 380bp length for 35 ingroup taxa (+ 5 outgroups)
- Total length aligned = **133,424bp**
 - **5.33 million total bp** of data
 - Contrast to Chen et al. 2013
 - **5 loci**, total length aligned **4,518bp**
 - **429,210 total bp**, 95 taxa



CLUPEIFORMES

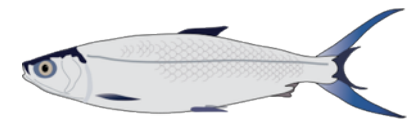
Engraulidae



Chirocentridae + Clupeidae

GONORHYNCHIFORMES

Gonorhynchidae



Chanidae

Phractolaemidae +

CYPRINIFORMES

Catosomidae

+
Cyprinidae



GYMNOTIFORMES

Apteronotidae + Sternopygidae



Hypopomidae

Gymnotidae

Citharinidae

+
Distichodontidae



CHARACIFORMES

Characidae + Ctenoluciidae



Erythrinidae

Hepsetidae +
Alestidae



Loricariidae

SILURIFORMES

Plotosidae + Ariidae



Ictaluridae +
Pangasidae

Malapteruridae + Mochokidae



Otocephala

Ostariophysi

Otophysi

Characiphysi

ML UCE Tree

CLUPEIFORMES

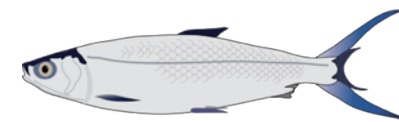
Engraulidae



Chirocentridae + Clupeidae

GONORHYNCHIFORMES

Gonorhynchidae



Chanidae

Phractolaemidae +

CYPRINIFORMES

Catosomidae

+
Cyprinidae



GYMNOTIFORMES

Apteronotidae + Sternopygidae



Hypopomidae

Gymnotidae

Citharinidae



+
Distichodontidae

Citharinoidei

CHARACIFORMES

Characidae + Ctenoluciidae



Erythrinidae

Hepsetidae +

Alestidae



Loricariidae

Plotosidae + Ariidae



Ictaluridae +

Pangasidae



Malapteruridae + Mochokidae

Otocephala

Ostariophysii

Otophysi

Characiphysi

100

100

100

100

100

100

100

100

100

100

100

100

100

100

29

100

100

100

100

100

Characoidei

97

100

97

100

100

100

100

100

100

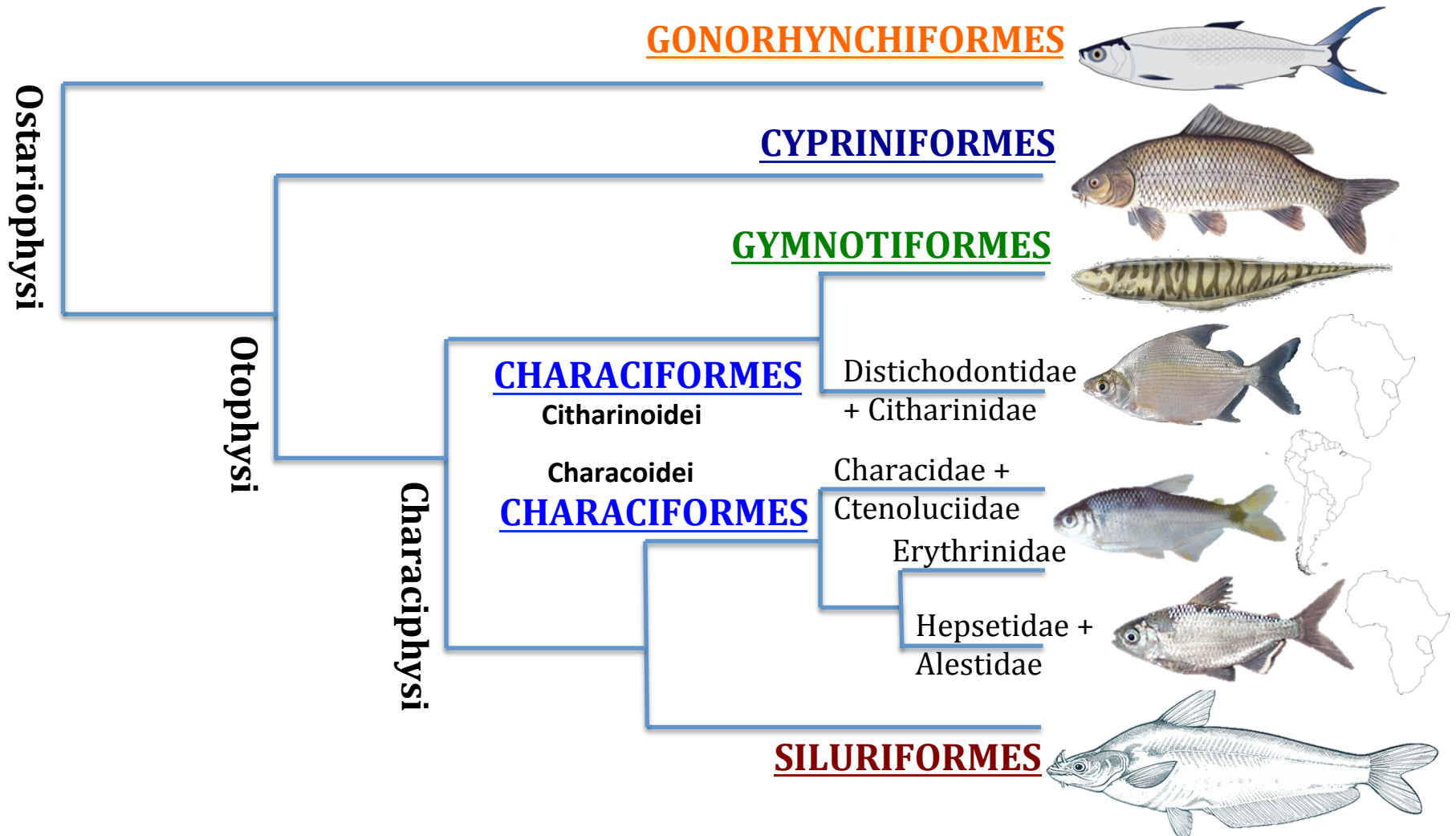
100

100

100

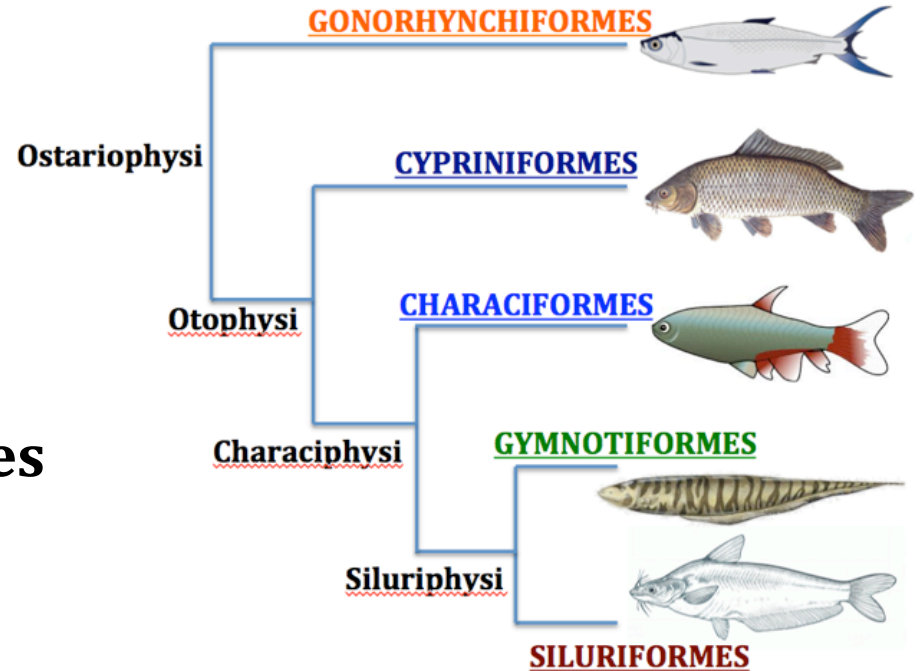
ML UCE Tree

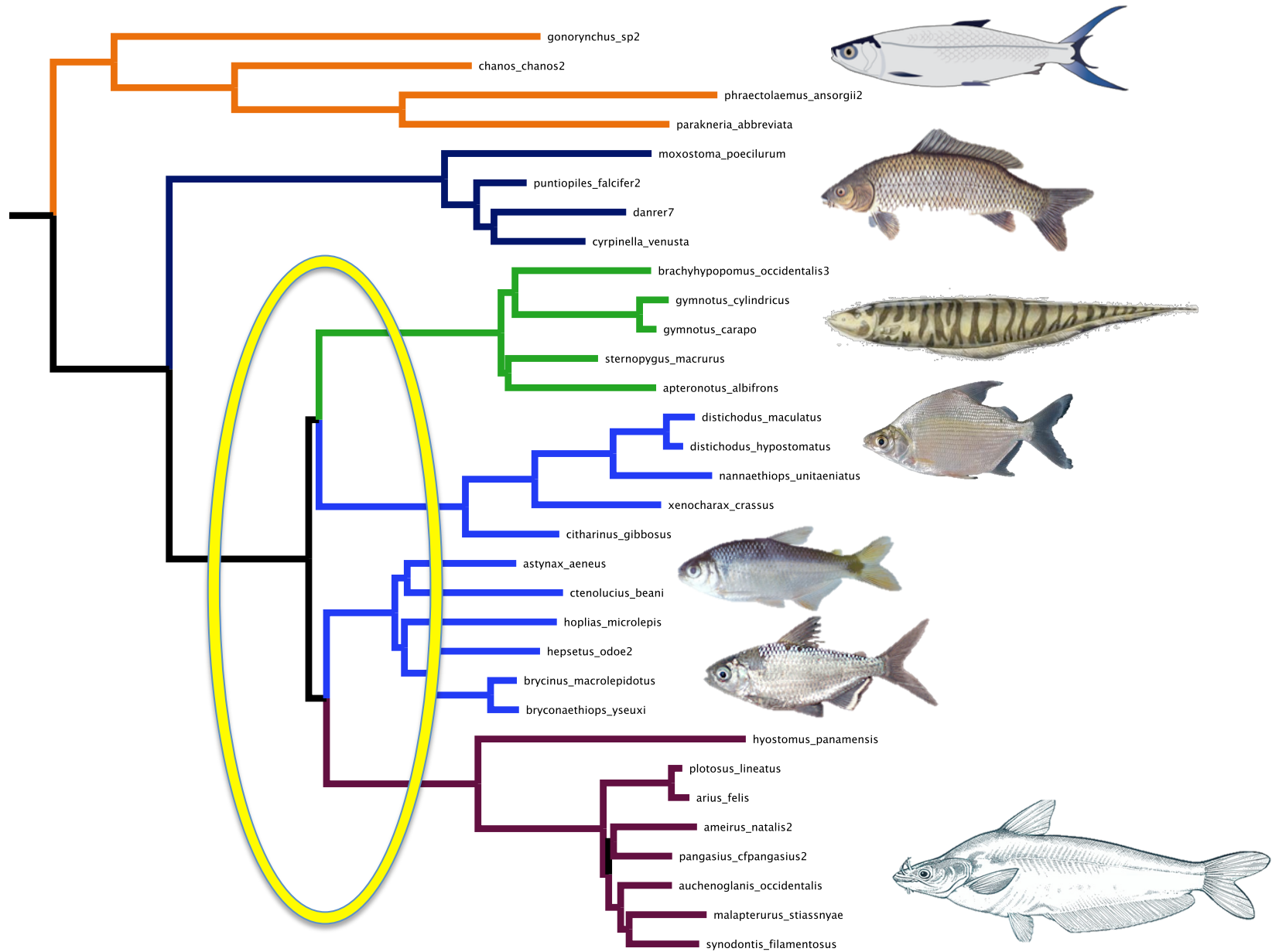
Ultraconserved Elements Phylogeny



Expected Relationships

- **Monophyletic Characiformes**
 - Similar morphologies, assumed deep homology, gestalt
- **Monophyly of Siluriphysi (Siluriformes + Gymnotiformes)**
 - Many neurological, sensory features related to electroreception that are thought to be homologous - difficult to explain as convergence (Fink and Fink, 1996)
 - Unique shared low frequency ampullary electroreception
 - Laminated electrosensory lateral line lobe

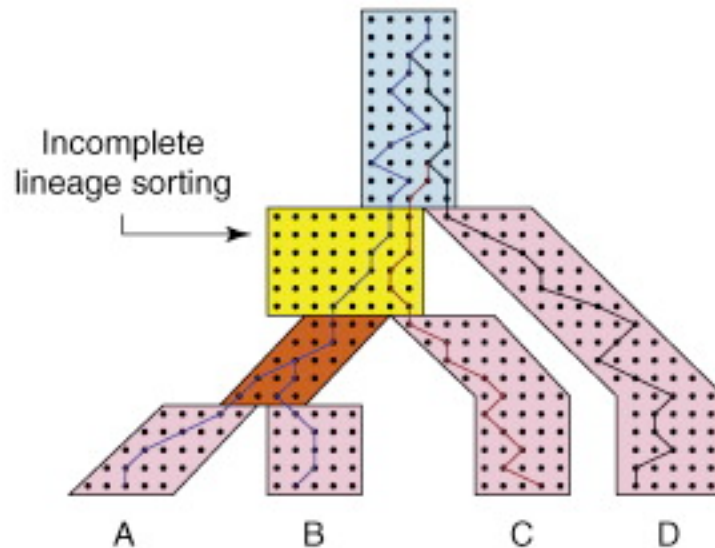




0.04

Potential Issues

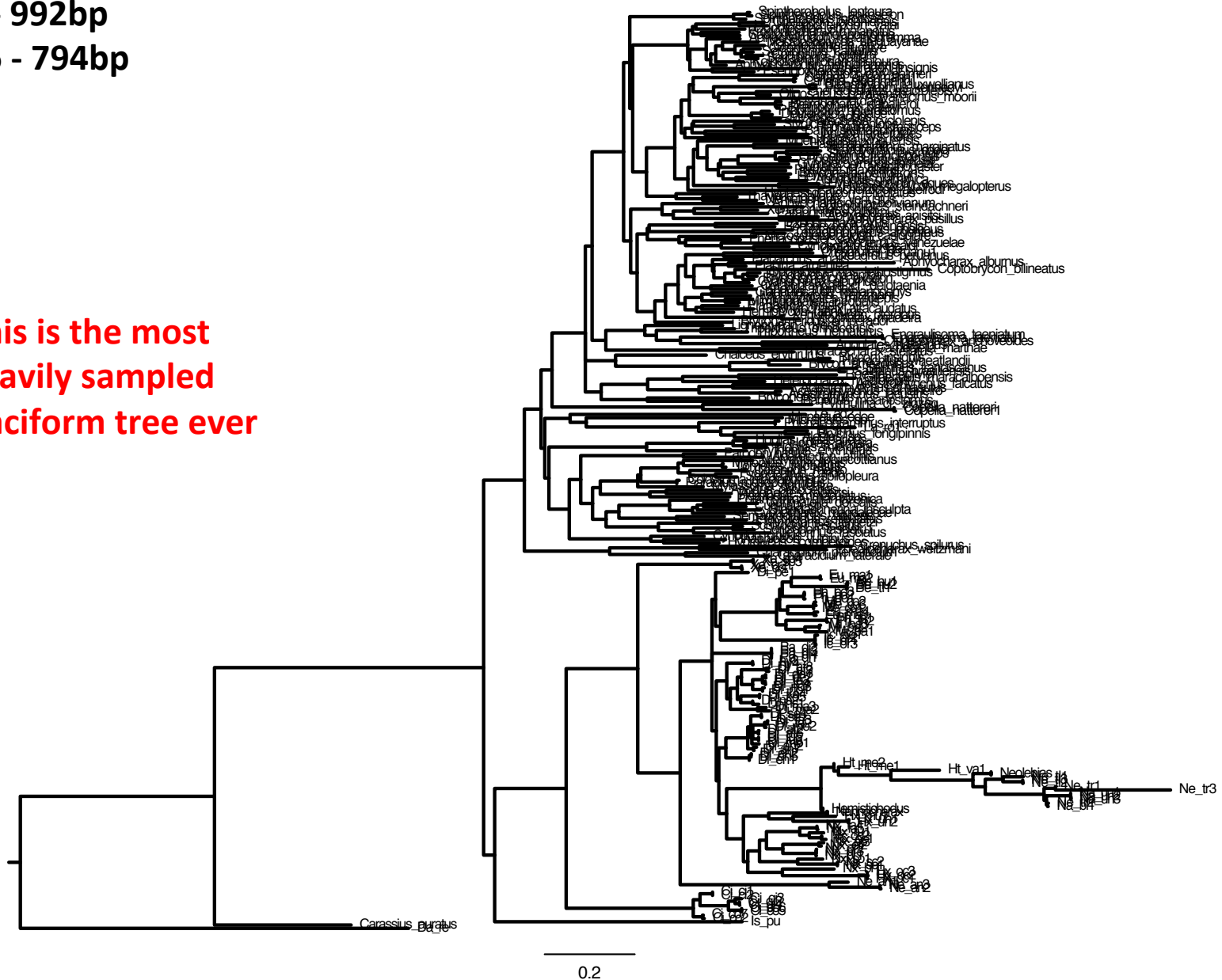
- 35 taxa representing 1/5 of all vertebrates
- Incomplete lineage sorting (Rapid radiation, large population sizes) – “mis-sorting”
 - need coalescent approach (working on Species Tree)
 - model gene tree discordance based on coalescent



282 Characiform Taxa Phylogeny (Combines Oliveira et al. 2011 w/ Arroyave & Stiassny)

cytb - 992bp
myh6 - 794bp

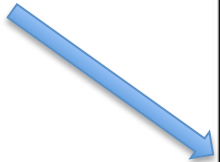
This is the most heavily sampled characiform tree ever



282 Characiform Taxa Phylogeny (Combines Oliveira et al. 2011 w/ Arroyave & Stiassny)

cytb - 992bp
myh6 - 794bp

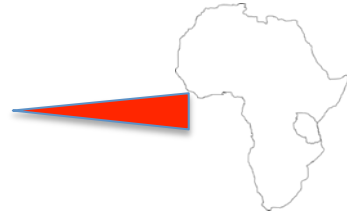
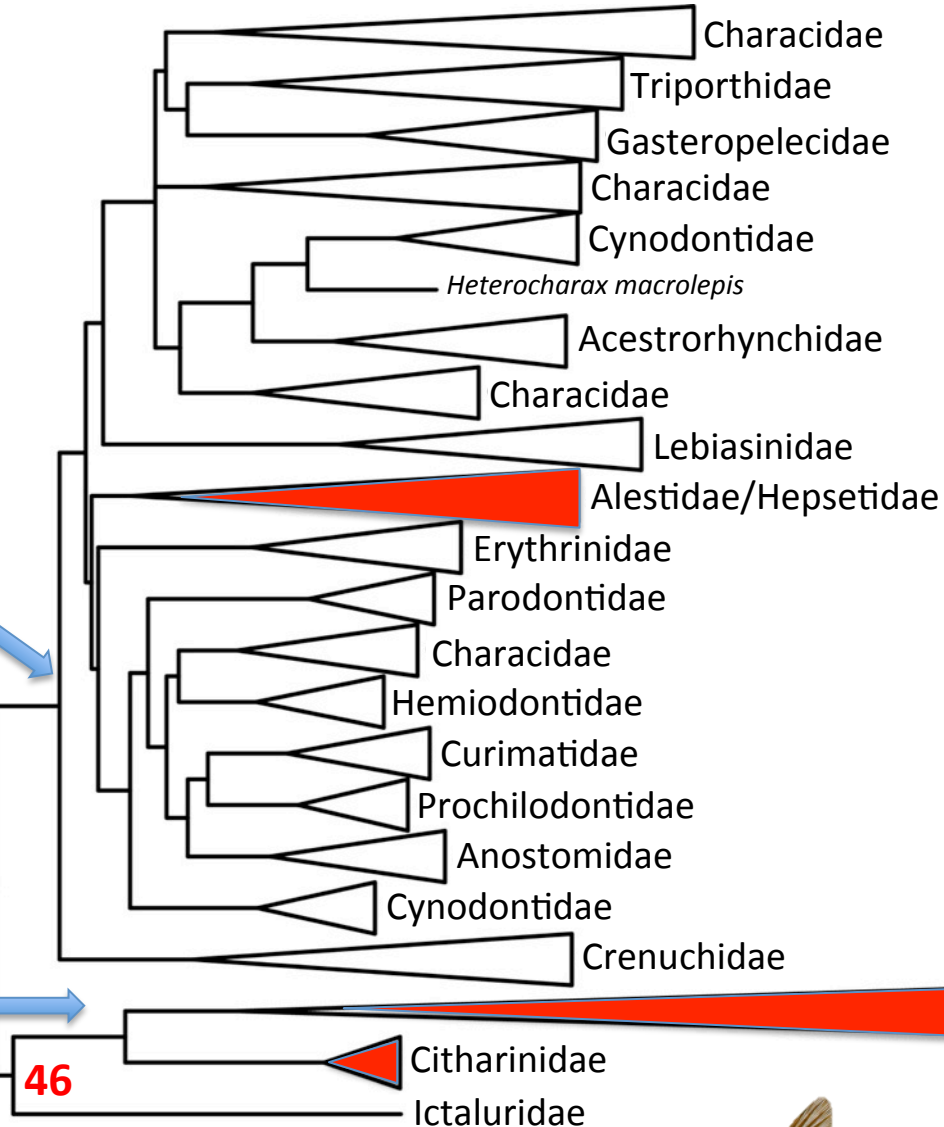
**Suborder
Characoidei**



**Suborder
Citharinoidei**

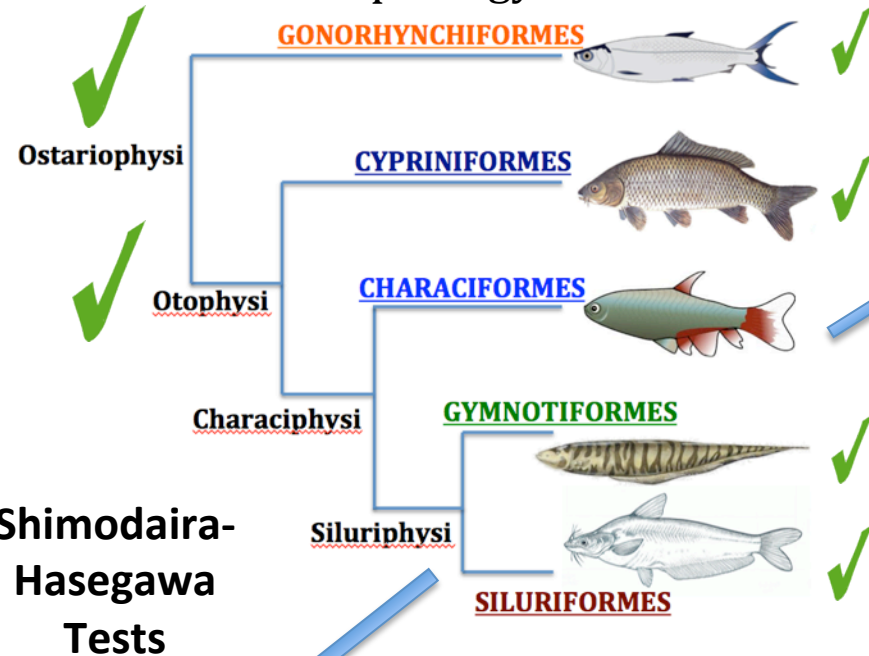


46



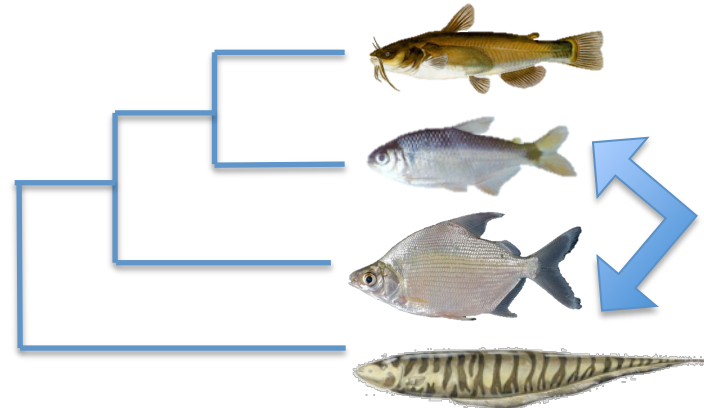
Previous Hypotheses

Fink and Fink Morphology



Non-monophyly of Characiformes
All molecular studies that include both suborders with more than one species represented e.g.
Chen et al. 2013, Nakatani et al. 2011

Non-monophyly of Siluriphysi
Never recovered in a molecular analysis



Relationships of Characiphysi

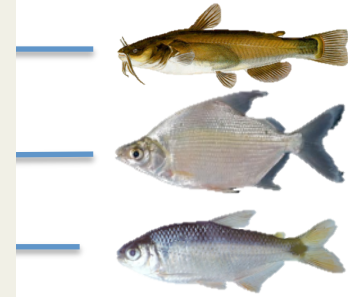
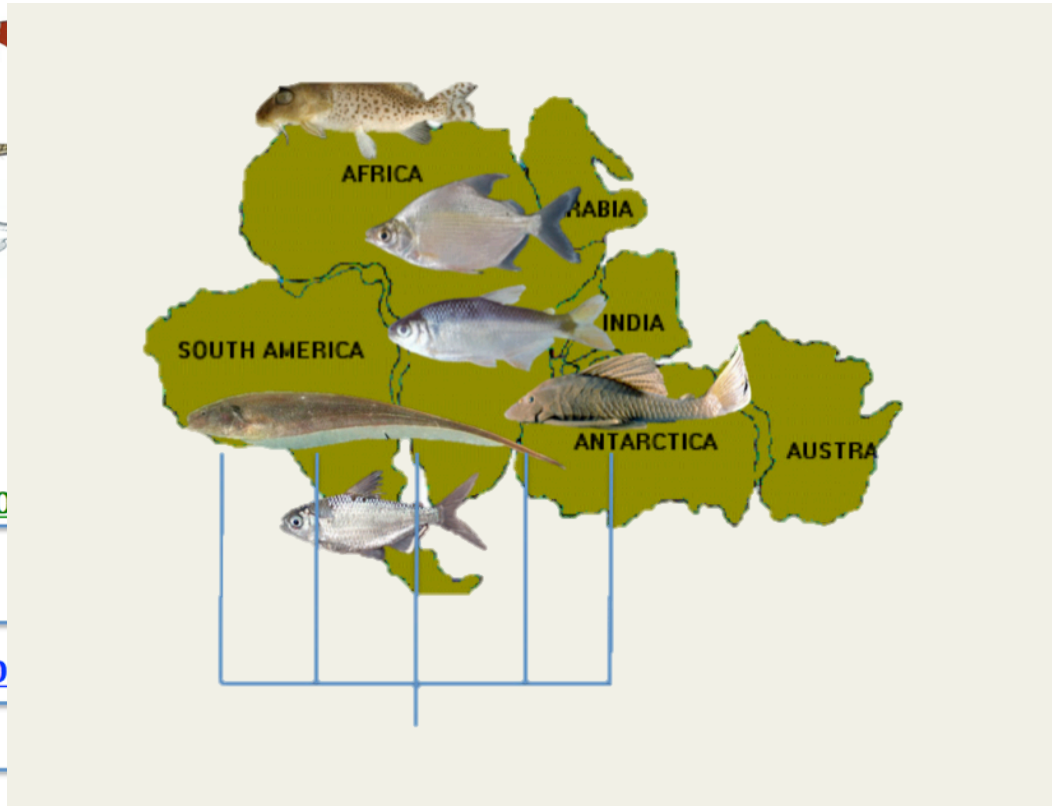
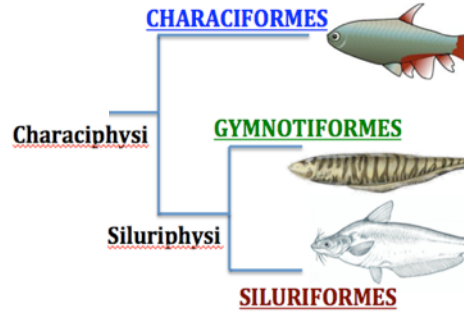
Fink and Fink Morphology

Chen et al. 2013, Nakatani et al. 2011

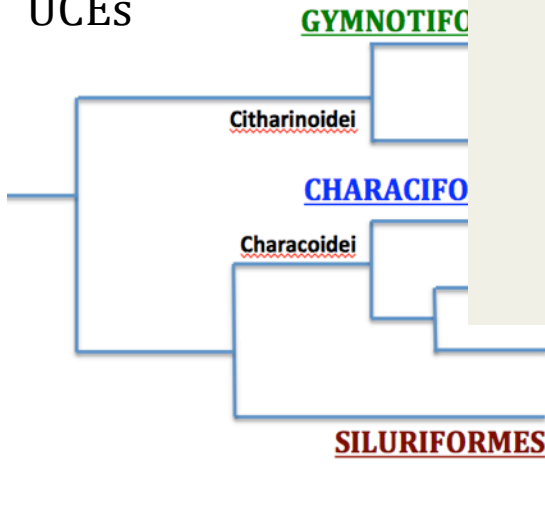
myh6 + cytb, Ortí & Meyer 1997

Siluriforms + Characoidei

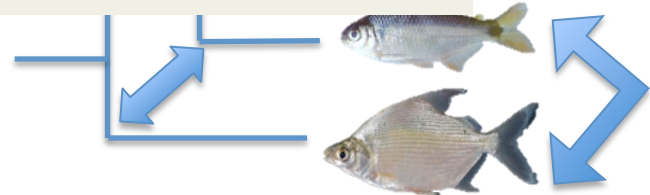
Siluriforms + Citharinoidei



UCEs



& Meyer 1996,
Lacour et al. 2013
Characiformes



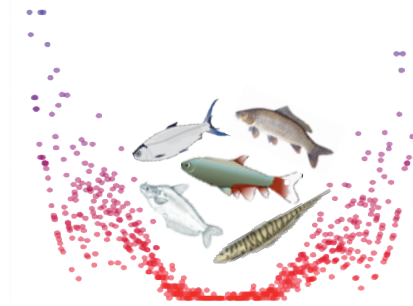
Conclusions

- Ultraconserved elements provide a powerful new way to sample hundreds of loci (millions of bp) from genomic fragments
- UCE phylogeny recovers novel relationships of Ostariophysi
 - Non-monophyly of Characiformes
 - Characoidei are sister to Siluriformes
 - Citharinoidei are sister to Gymnotiformes
- Relationships among Characiphysi are difficult to tease out even with lots (and lots) of data
 - Rapid radiations still hard to resolve, even 150my later



Take Home Questions

- Molecular analyses are not giving us a consistent answer about the relationships of Ostariophysi
 - Why?
- Genomic data will not necessarily give us the true tree; however, UCEs are giving us additional and novel information.
 - Why? Still learning.



Acknowledgments

Bill Ludt, Mike Harvey, Brian Smith at the Louisiana State University Museum of Natural Science

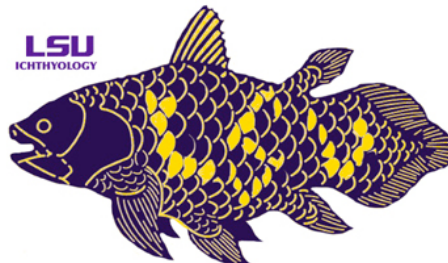
<http://ultraconserved.org>



Caleb McMahan, Bill Fink, James Albert, Jairo Arroyave, Melanie Stiasny, Brant Faircloth, Mike Alfaro



@LSU_FISH

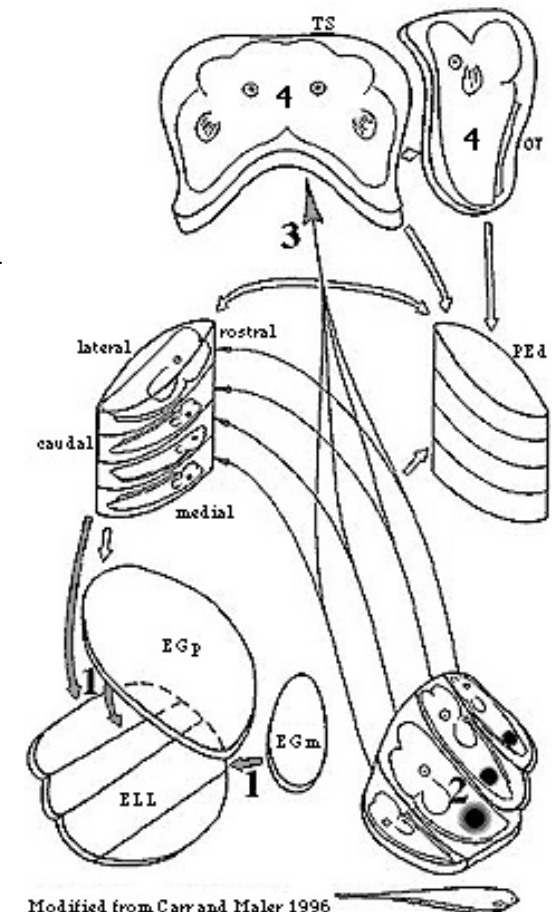




ありがとう

Deep similarities in the nervous systems of Siluriforms & Gymnotiforms

- described in characters 1-15 of Fink and Fink(1996: 219-220).
- also Albert and Crampton, 2006
- figure Carr and Maler 1986 showing the four parallel pathways of electrosensory information through the brain of *Eigenmannia*. The medial pathway is the ampullary organ path shared with catfishes.
- Low frequency, ampullary-shaped electroreceptor organs tuned to low frequency (less than 30hz)



Characiformes monophyly

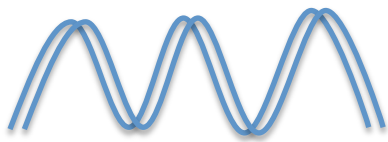
- Adipose fin, well developed teeth, decurved lateral line, unspecialized pharyngeal teeth, gap between hypural and centrum...



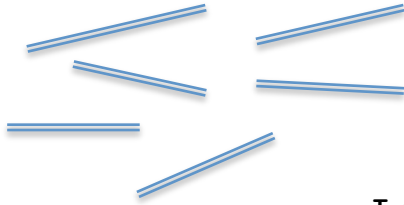
TABLE 12.2**Diversity of Electric Fish Taxa and Salient Features of Electrogenic Systems.**

Electrogenic Taxon	Common Name(s)	Higher Taxon	Spp.	Electric Organ(s)	EOD	Use	Voltage	Habitat	Refs
Torpedinidae	Electric rays	Torpediniformes	50	Branchial/axial m.	Wave	Interm.	weak, strong	Mar.	—
Rajidae	Skates	Rajiformes	228	Axial m. (tail)	Wave	Interm.	weak	Mar.	119
Gymnarchidae	Gymnarchus	Mormyriiformes	1	Axial m.	Wave	Contin.	weak	FW	155
Mormyridae	African electric fishes	Mormyriiformes	234	Axial m.	Pulse	Contin.	weak	FW	209
Auchenoglanis	Giraffe catfishes	Siluriformes	≥1	NA	?	Interm.	weak	FW	18
Clarias	Walking catfishes	Siluriformes	≥1	NA	Pulse	Interm.	weak	FW	17
Ictalurus	Bullhead catfishes	Siluriformes	≥2	NA	DC	Interm.	weak	FW	182
Malapterurus	Electric catfishes	Siluriformes	12	Pectoral m.	Pulse	Interm.	strong	FW	155
Ompok	Sheatfish	Siluriformes	≥1	NA	?	Interm.	weak	FW	162
Synodontis	Upside-down catfishes	Siluriformes	≥3	Swim-bladder m.	Pulse	Interm.	weak	FW	101
Gymnotidae	Gymnotids/electric eels	Gymnotiformes	35	Hypaxial m.	Pulse	Contin.	weak, strong	FW	5
Hypopomidae	Bluntnose/pintail knifefishes	Gymnotiformes	29	Hypaxial m.	Pulse	Contin.	weak	FW	5
Rhamphichthyid	Sand knifefishes	Gymnotiformes	13	Hypaxial m.	Pulse	Contin.	weak	FW	5
Sternopygidae	Glass & rattail knifefishes	Gymnotiformes	39	Pterygiophore/epaxial m.	Wave	Contin.	weak	FW	5
Apteronotidae	Ghost knifefishes	Gymnotiformes	60	Spinal motoneurons	Wave	Contin.	weak	FW	5
Astroscopus	Stargazers	Uranoscopidae	4	Extrinsic eyeball m.	Pulse	Interm.	strong	Mar.	16

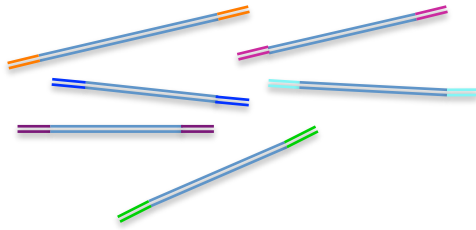
Abbreviations: contin., continuous or with transient cessation; DC, direct current; interm., intermittent, used only during courtship and aggressive displays; Spp., species-richness estimates; EOD, electric organ discharge; Mar, Marine; FW, Fresh water; Refs, references. Electric organ(s) refers to adult condition; all derived from muscles except the neurogenic organ of Apteronotidae. Voltage: weak <1V, strong >10V.



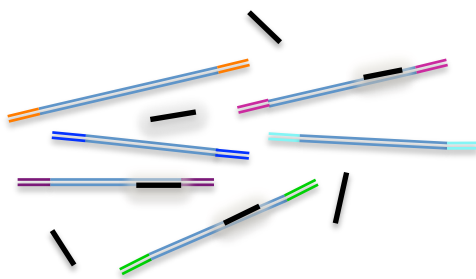
Fragment
Genomic DNA



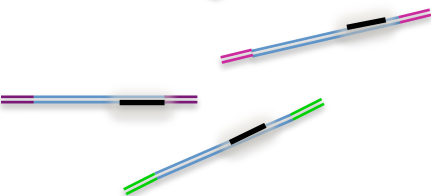
Ligate Barcoded
Adapters



Hybridize to
RNA Probes



Enrich for
Hybridized
Fragments



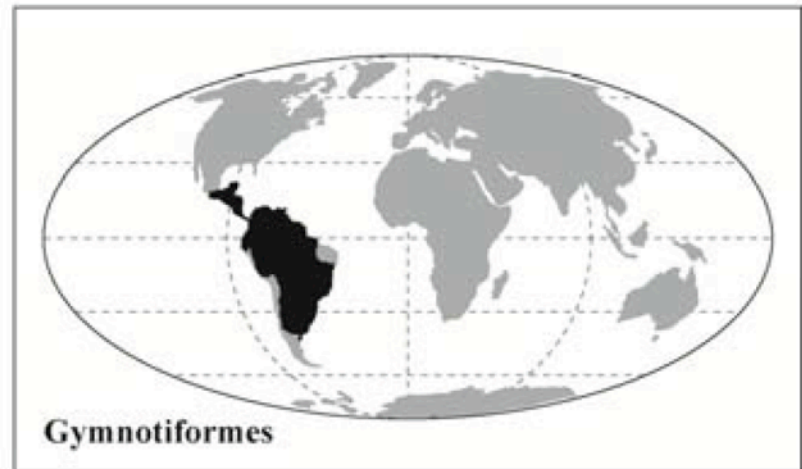
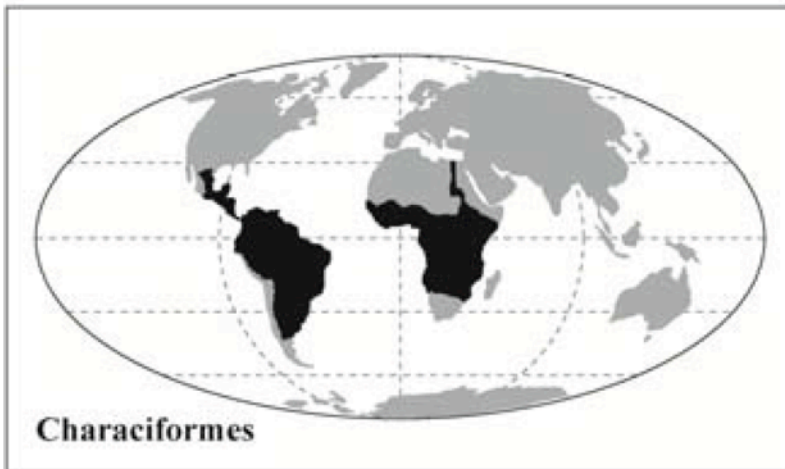
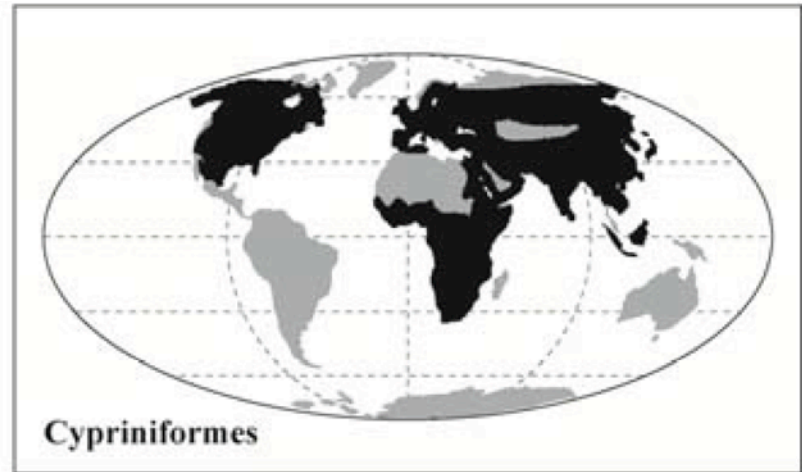
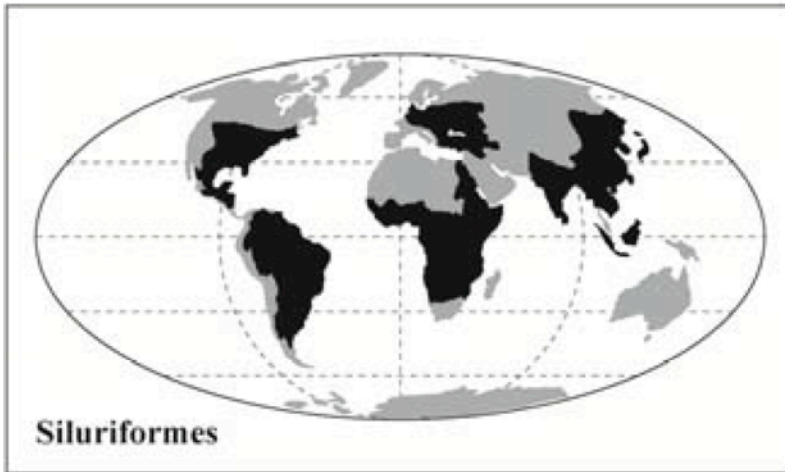
1. Genomic DNA sheared into appropriately sized fragments

2. Adaptors with barcodes unique to each individual added; PCR amplification.

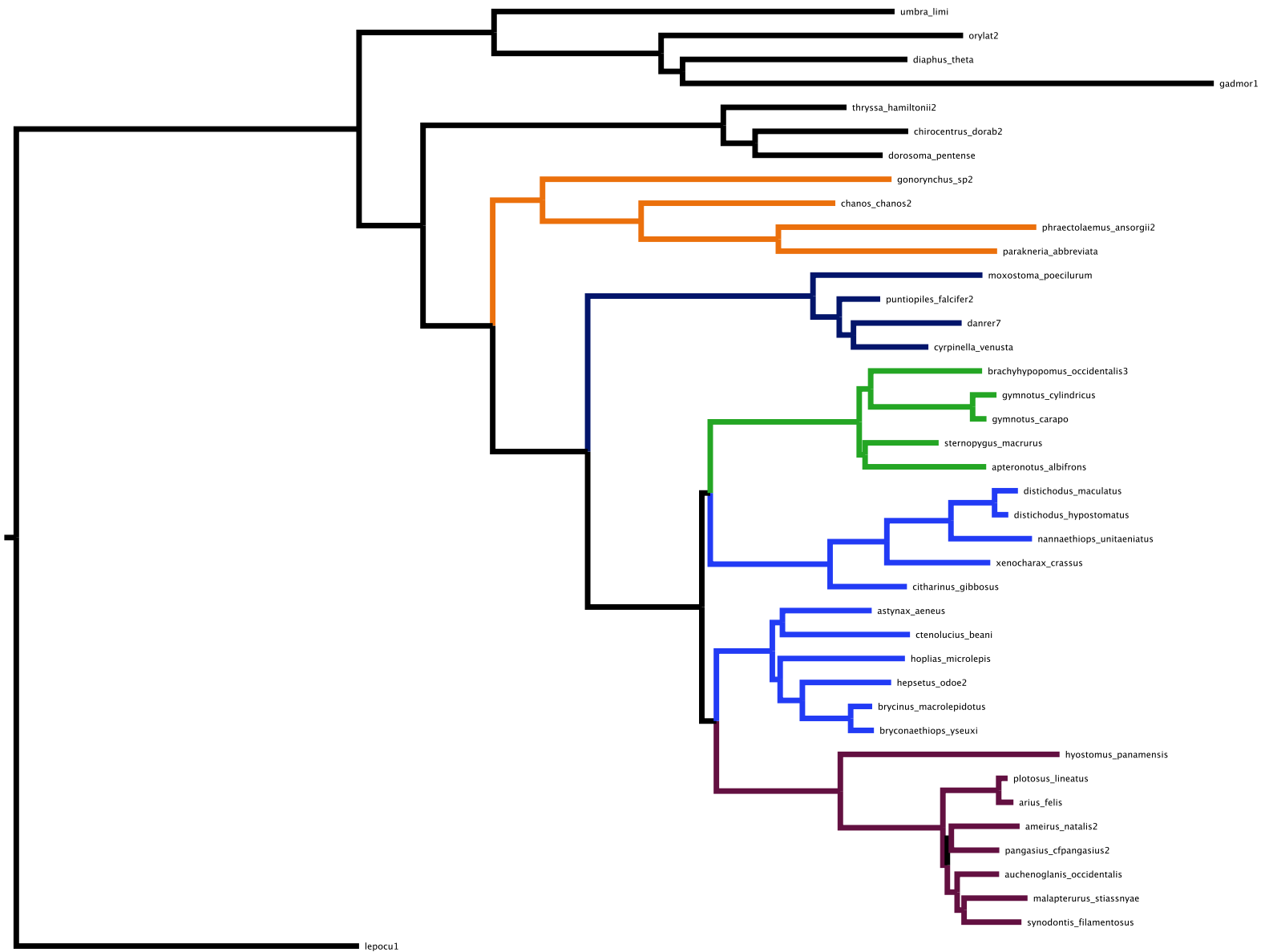
3. RNA probes designed from UCEs mixed with genomic fragments; probes hybridize to select fragments from corresponding loci.

4. Enrichment pulls out only those fragments with annealed probes; PCR amplification; pooling of samples for sequencing.

5. Next generation sequencing on Illumina GAIIx yields many reads for each UCE.



- From Chen et al 2013



0.04

SH Tests

- All four of the constraints produce trees that are much worse than the best tree. D(LH) is the difference in log likelihoods between best and any of the other trees.
- Model optimization, best Tree: -758991.212395
Found 4 trees in File constrained_best.phy
 - Monophyletic characiforms - Tree: 0 Likelihood: -759081.103240 D(LH): -89.890845 SD: 26.490494 Significantly Worse: Yes (5%), Yes (2%), Yes (1%)
 - catfish + knifefish - Tree: 1 Likelihood: -759077.714547 D(LH): -86.502152 SD: 27.735603 Significantly Worse: Yes (5%), Yes (2%), Yes (1%)
 - characiforms (catfish + knifefish) Tree: 2 Likelihood: -759099.910817 D(LH): -108.698422 SD: 30.615873 Significantly Worse: Yes (5%), Yes (2%), Yes (1%)
 - cyprinids sister to characiforms (catfish + knifefish) Tree: 3 Likelihood: -759099.910816 D(LH): -108.698 421 SD: 30.615958 Significantly Worse: Yes (5%), Yes (2%), Yes (1%)

Results UCE

-
- Lengths
- -----
- Total length(aln) 133424
- Total characters(aln) 132916
- 95 CI length(aln) 32.3785943853
- Minimum avg. ambiguities(aln) 0.0
- Maximum avg. ambiguities(aln) 0.0
- Percent avg. ambiguities(aln) as fxn of mean(length) 0.0
- Percent avg. ambiguities(aln) as fxn of mean(length) 0.0
-
- Taxa
- -----
- Average(taxa) 33.2991452991
- 95 CI(taxa) 0.229040019554
- min(taxa) 30
- max(taxa) 39
- Count(taxa:# alns) {32: 47, 33: 66, 34: 41, 35: 48, 36: 35, 37: 18, 38: 8, 39: 2, 30: 42, 31: 44}
-
- Base composition
- -----
- Bases {'A': 824190, 'C': 650388, 'G': 641930, '-': 1243408, 'T': 829778, '?': 289103}
- Sum(all) 4478797
- Sum(nucleotide only) 2946286
- Missing data from trim (%) 0.0