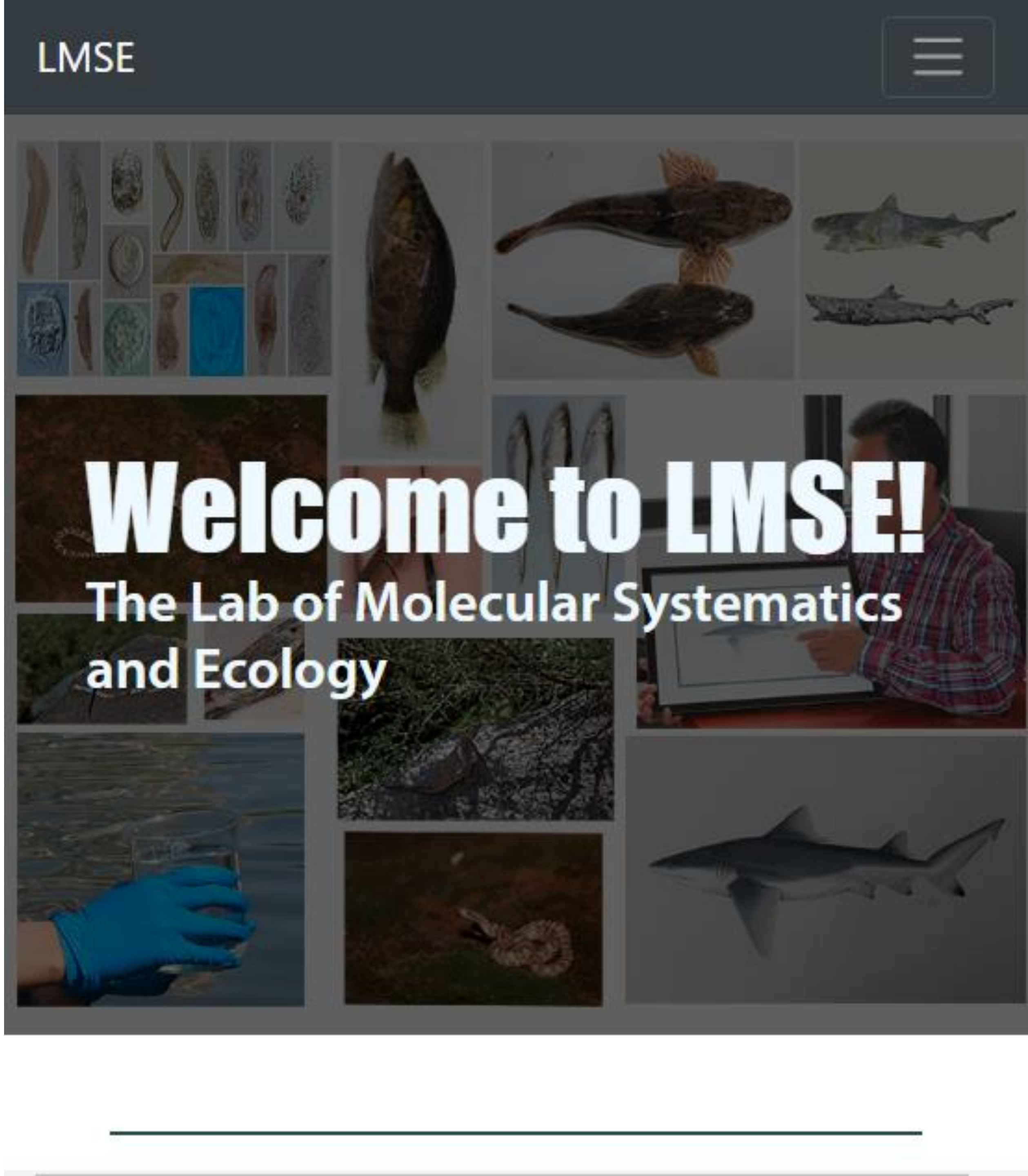
Previously …

Our Lab,
Main experiment methods

- Make use of EvolMarkers tool to design baits sequences

- Experimental procedure:
    construct gene library
    doing gene capture

- Analysis data

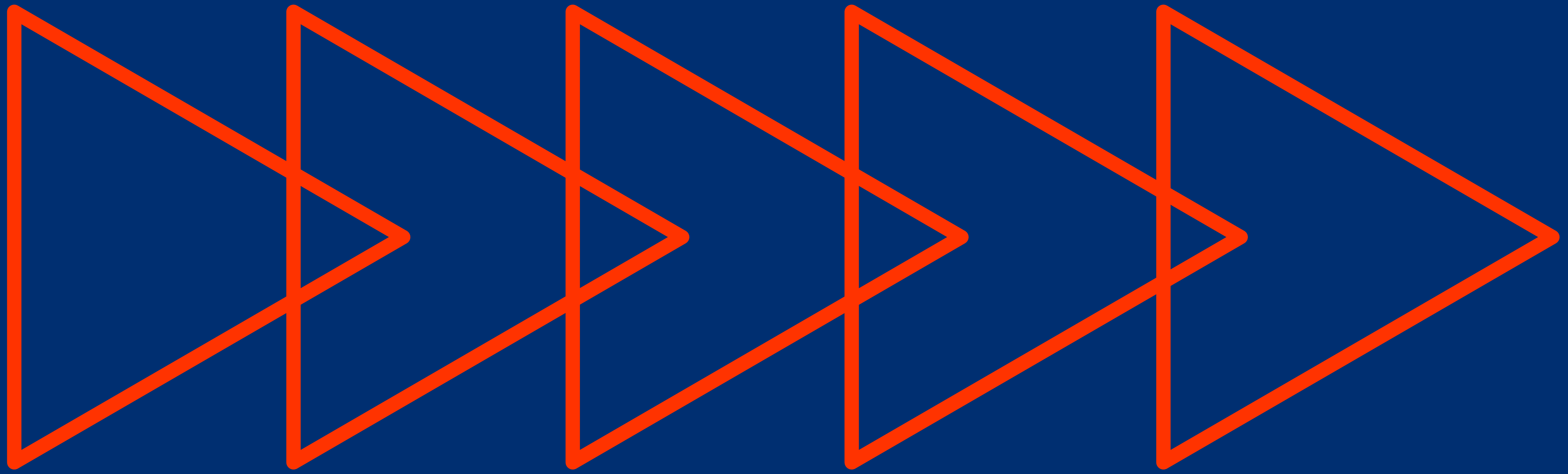


LMSE

Welcome to LMSE!

The Lab of Molecular Systematics and Ecology

	Jan 7 (Mon.)	Jan 8 (Tue.)	Jan 9 (Wend.)	Jan 10 (Thur.)	Jan 11 (Fri.)
8:30 – 9:00 **9:00 – 9:15 discussion**	EvolMarkers2 Junman Huang	Date filtering, partition Hao Yuan	Population structure, AMOVA, PCA Qingwen Xue	F-dist, Bayescan, Other adaptive methods Longlong & Suhan	Intro AI Liang Lu
			Tea Break		
9:30 – 10:00 **10:00 – 10:15 discussion**	Lib prep & gene cap Lifang Peng	Gene tree, species tree Guoxin Yin	Spatial structure, population dynamics Huirui & Ying	GWAS introduction, improved method Ziqiang Gong	Convolutional network Liang Lu
			Tea Break		
10:30 – 11:00 **11:00 – 11:15 discussion**	Read assembling Junman Huang	Time calibration, topology test Guoxin Yin	Species delimitation Lei & Songjun	Environment GWAS, pedigree deducing Ziqiang Gong	GANs Hao Yuan
			Lunch		
1:30 – 2:00 **2:00 – 2:15 discussion**	Post assembling data processing Hao Yuan	Biogeography, character mapping Yinyi Yang	ABC Anirban Sarker	Transcriptomic analysis Tao Zhou	Protein folding Liang Lu
			Tea Break		
2:30 – 3:00 **3:00 – 3:15 discussion**	Molecular evolution Chenhong Li	SNP calling SNPs vs sequences Qiaoyun Ai	Fastsimcoal2 Lifang Peng	Comparative genomics, EP Hao Yuan	Genome prediction Hao Yuan
			Tea Break		
3:30 – 4:00 **4:00 – 4:15 discussion**	Population genetics Chenhong Li	Summary statistics, Arlequin Qiaoyun Ai	Land markers Qiaoyun Ai	Open: How to identify phenotype associated genes	Open: Ideas applying AI



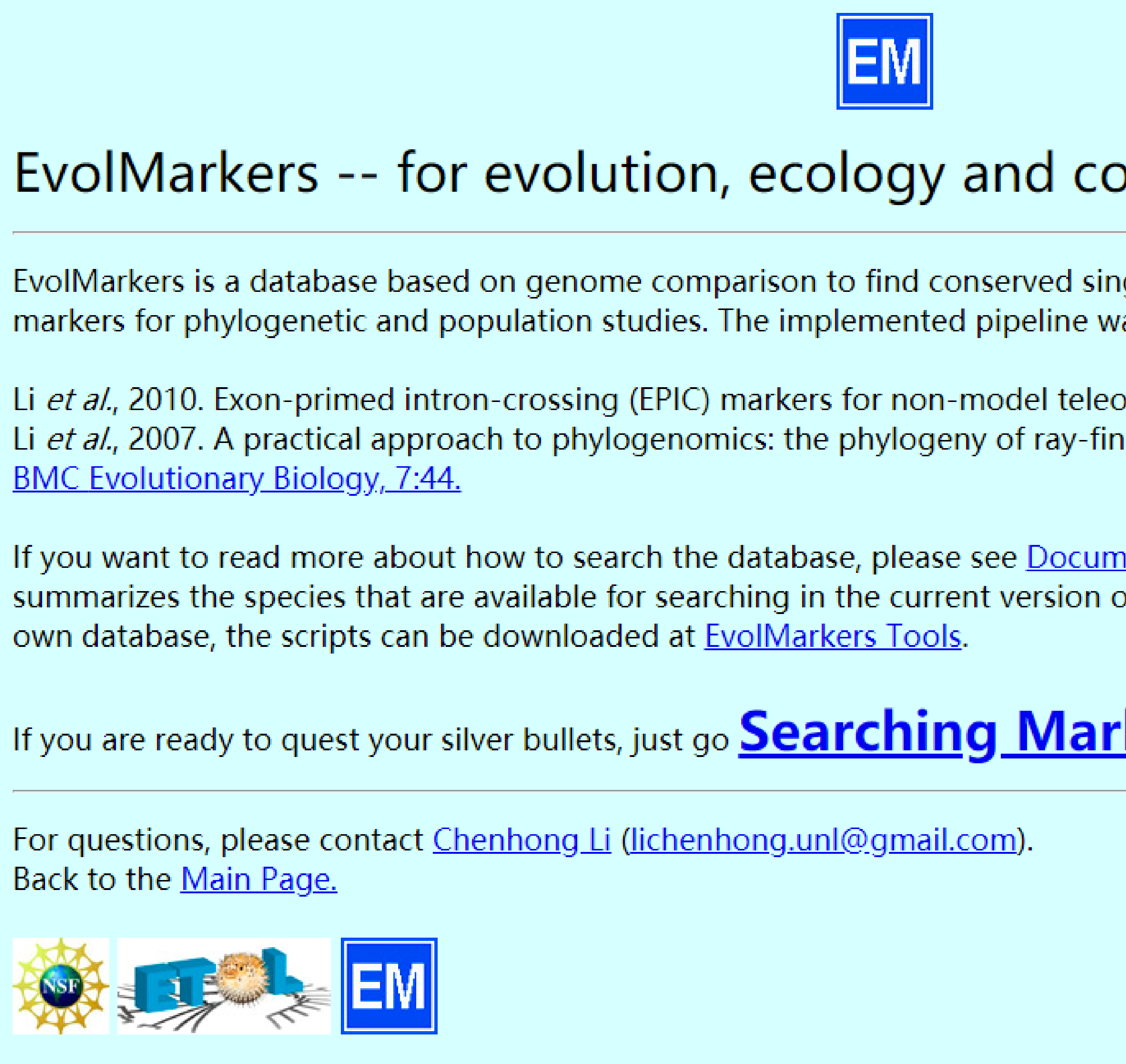
GOTA MOVE ON ...

EvolMarkers 2

EvolMarkers,
Online Tool

For evolution, ecology and conservation studies

Website link:
http://bioinformatics.unl.edu/cli/evolmarkers/



EvolMarkers -- for evolution, ecology and co

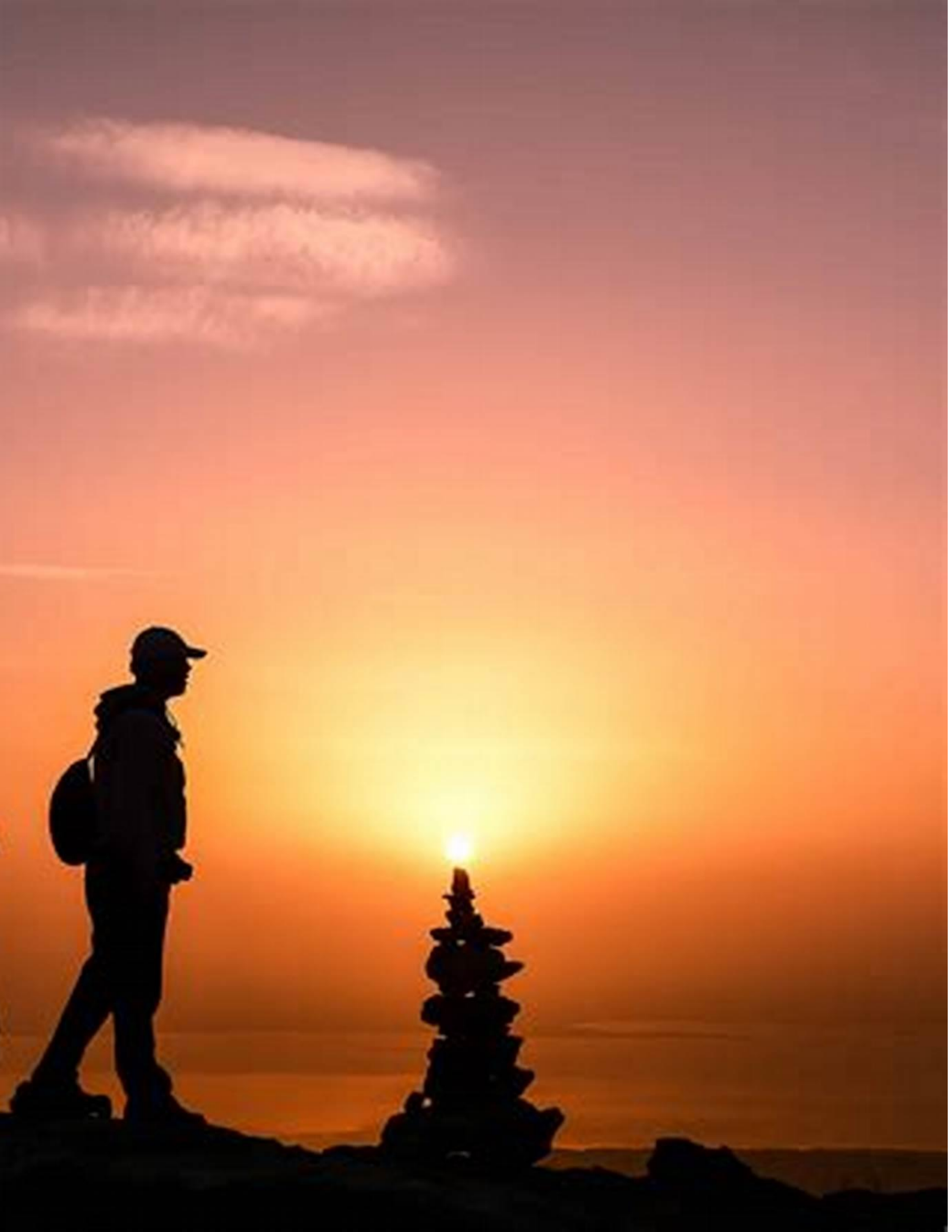
EvolMarkers is a database based on genome comparison to find conserved sin
markers for phylogenetic and population studies. The implemented pipeline wa

Li *et al.*, 2010. Exon-primed intron-crossing (EPIC) markers for non-model teleo
Li *et al.*, 2007. A practical approach to phylogenomics: the phylogeny of ray-fin
BMC Evolutionary Biology, 7:44.

If you want to read more about how to search the database, please see Docum
summarizes the species that are available for searching in the current version o
own database, the scripts can be downloaded at EvolMarkers Tools.

If you are ready to quest your silver bullets, just go **Searching Mar**

For questions, please contact Chenhong Li (lichenhong.unl@gmail.com).
Back to the Main Page.



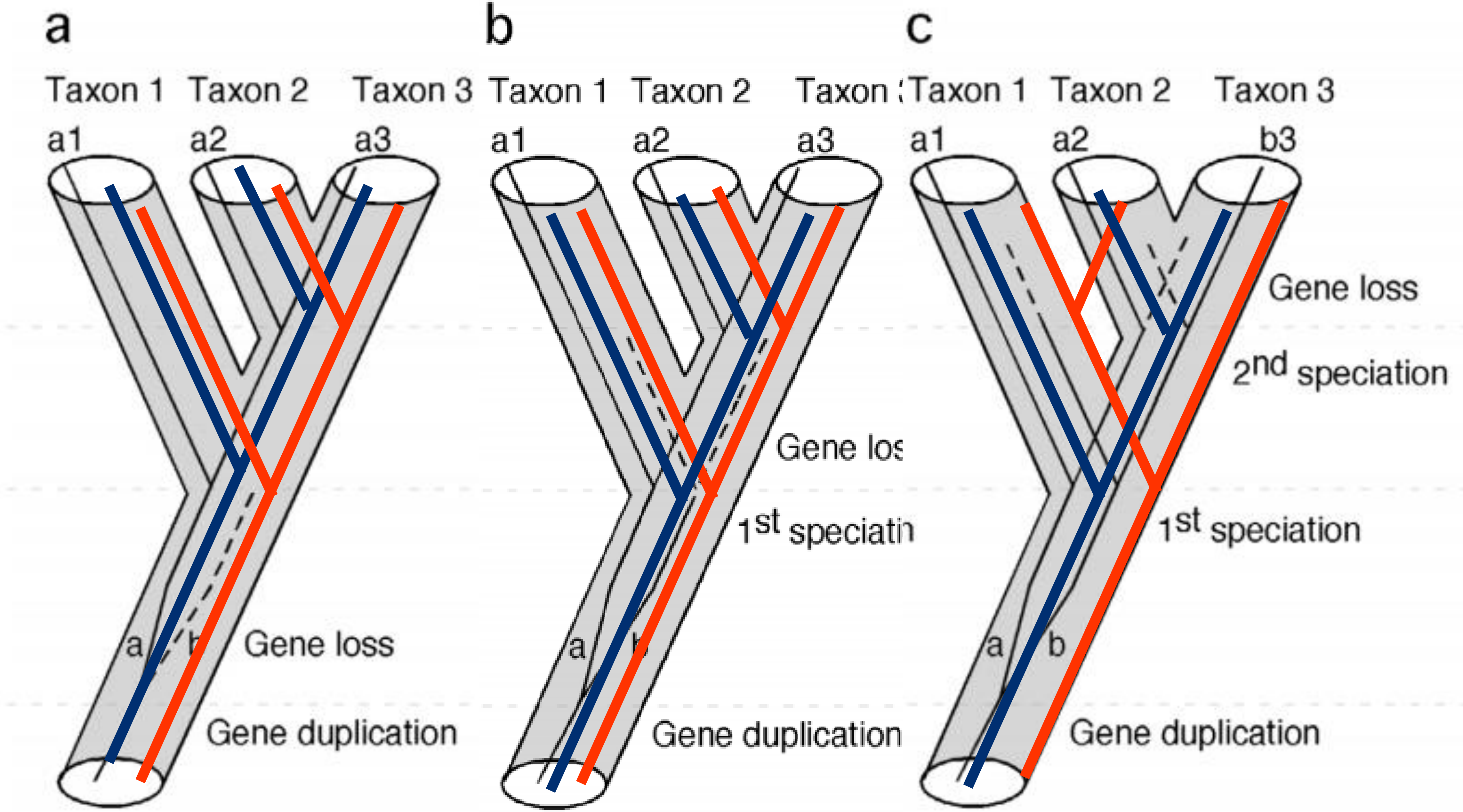
What can we do, in EvolMarkers ?

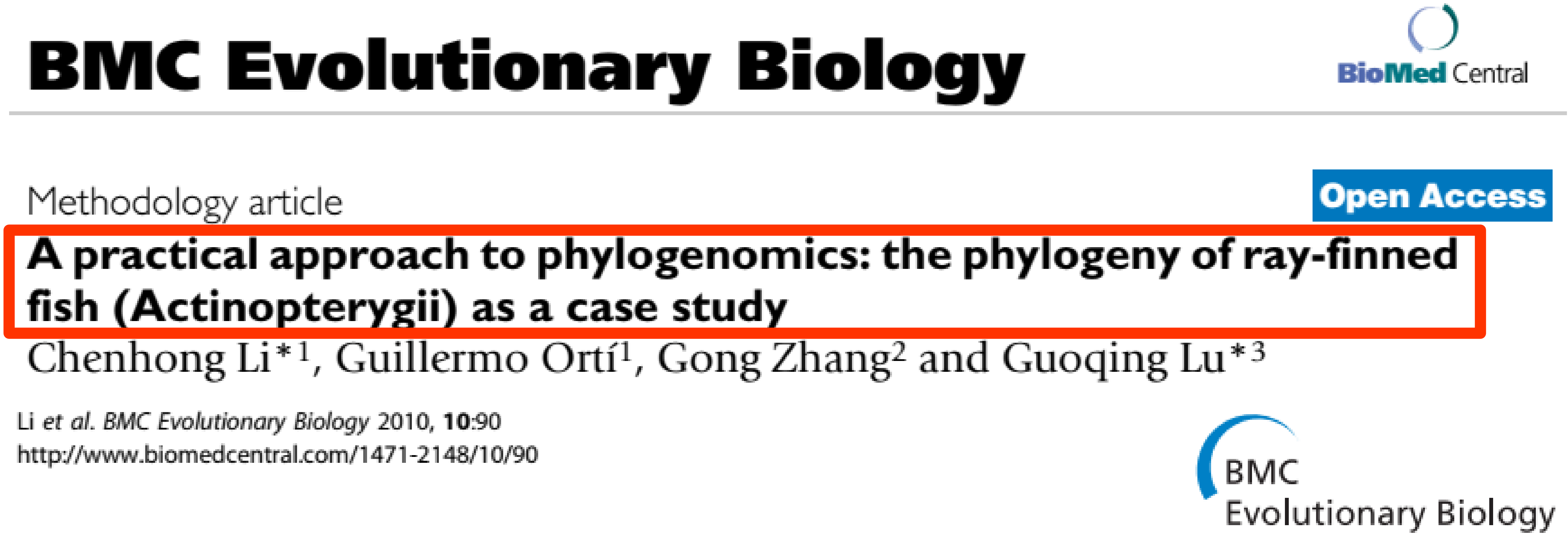
- **Identify single-copy coding markers through genome comparison**
- Identify EPIC
  --- exon primed intron crossing
- Obtaining more raw data for analysis

……

“Why choose single-copy coding sequences? Why not choose multi-copy coding sequences?”

……

a
Taxon 1
Taxon 2
Taxon 3
a1
a2
a3
a
b
Gene loss
Gene duplication
b
Taxon 1
Taxon 2
Taxon
a1
a2
a3
Gene los
1st speciatin
a
b
Gene duplication
c
Taxon 1
Taxon 2
Taxon 3
a1
a2
b3
Gene loss
2nd speciation
1st speciation
a
b
Gene duplication



BMC Evolutionary Biology

BioMed Central

Methodology article

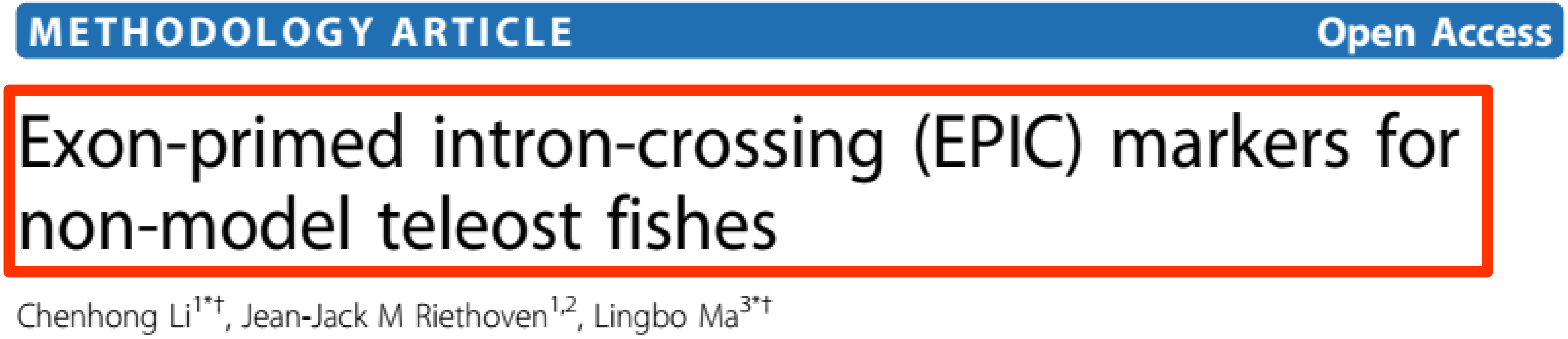
Open Access

A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study

Chenhong Li*[1], Guillermo Ortí[1], Gong Zhang[2] and Guoqing Lu*[3]

Li *et al. BMC Evolutionary Biology* 2010, **10**:90
http://www.biomedcentral.com/1471-2148/10/90

BMC
Evolutionary Biology



METHODOLOGY ARTICLE **Open Access**

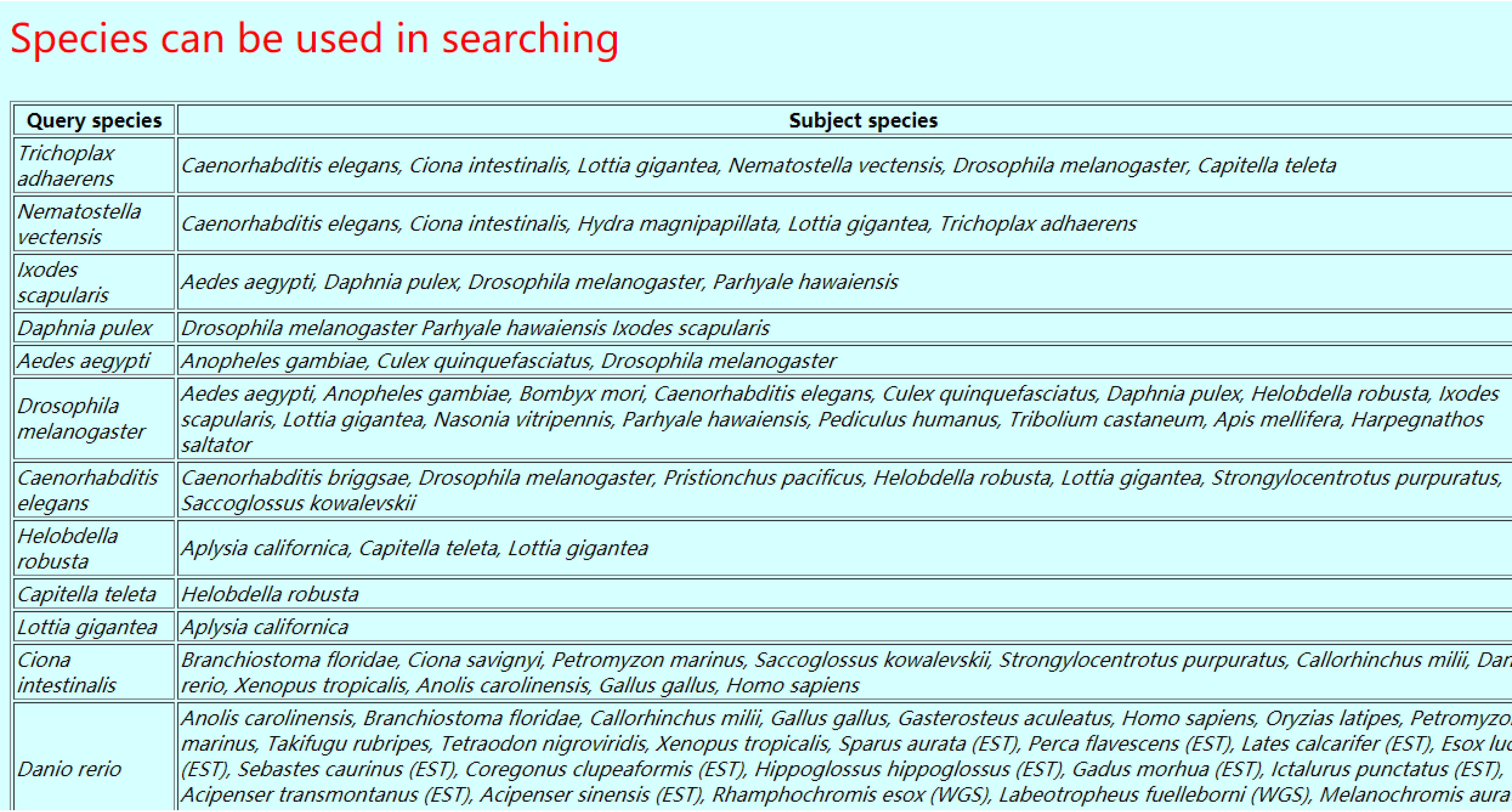
Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes

Chenhong Li[1*†], Jean-Jack M Riethoven[1,2], Lingbo Ma[3*†]

UPGRADE EVOLMARKERS 2

Compared with EvolMarkers 1, Shining point

- More query taxa
- Optimize markers
- Lively



Species can be used in searching

Query species	Subject species
Trichoplax adhaerens	*Caenorhabditis elegans, Ciona intestinalis, Lottia gigantea, Nematostella vectensis, Drosophila melanogaster, Capitella teleta*
Nematostella vectensis	*Caenorhabditis elegans, Ciona intestinalis, Hydra magnipapillata, Lottia gigantea, Trichoplax adhaerens*
Ixodes scapularis	*Aedes aegypti, Daphnia pulex, Drosophila melanogaster, Parhyale hawaiensis*
Daphnia pulex	*Drosophila melanogaster Parhyale hawaiensis Ixodes scapularis*
Aedes aegypti	*Anopheles gambiae, Culex quinquefasciatus, Drosophila melanogaster*
Drosophila melanogaster	*Aedes aegypti, Anopheles gambiae, Bombyx mori, Caenorhabditis elegans, Culex quinquefasciatus, Daphnia pulex, Helobdella robusta, Ixodes scapularis, Lottia gigantea, Nasonia vitripennis, Parhyale hawaiensis, Pediculus humanus, Tribolium castaneum, Apis mellifera, Harpegnathos saltator*
Caenorhabditis elegans	*Caenorhabditis briggsae, Drosophila melanogaster, Pristionchus pacificus, Helobdella robusta, Lottia gigantea, Strongylocentrotus purpuratus, Saccoglossus kowalevskii*
Helobdella robusta	*Aplysia californica, Capitella teleta, Lottia gigantea*
Capitella teleta	*Helobdella robusta*
Lottia gigantea	*Aplysia californica*
Ciona intestinalis	*Branchiostoma floridae, Ciona savignyi, Petromyzon marinus, Saccoglossus kowalevskii, Strongylocentrotus purpuratus, Callorhinchus milii, Dar rerio, Xenopus tropicalis, Anolis carolinensis, Gallus gallus, Homo sapiens*
Danio rerio	*Anolis carolinensis, Branchiostoma floridae, Callorhinchus milii, Gallus gallus, Gasterosteus aculeatus, Homo sapiens, Oryzias latipes, Petromyzo marinus, Takifugu rubripes, Tetraodon nigroviridis, Xenopus tropicalis, Sparus aurata (EST), Perca flavescens (EST), Lates calcarifer (EST), Esox luc (EST), Sebastes caurinus (EST), Coregonus clupeaformis (EST), Hippoglossus hippoglossus (EST), Gadus morhua (EST), Ictalurus punctatus (EST), Acipenser transmontanus (EST), Acipenser sinensis (EST), Rhamphochromis esox (WGS), Labeotropheus fuelleborni (WGS), Melanochromis aura*

volMarkers -- for evolution, ecology and conservation studies

volMarkers is a database based on genome comparison to find conserved single-copy exon (CDS) and intron (EPIC) markers for phylogenetic and opulation studies. The implemented pipeline was originally described by:

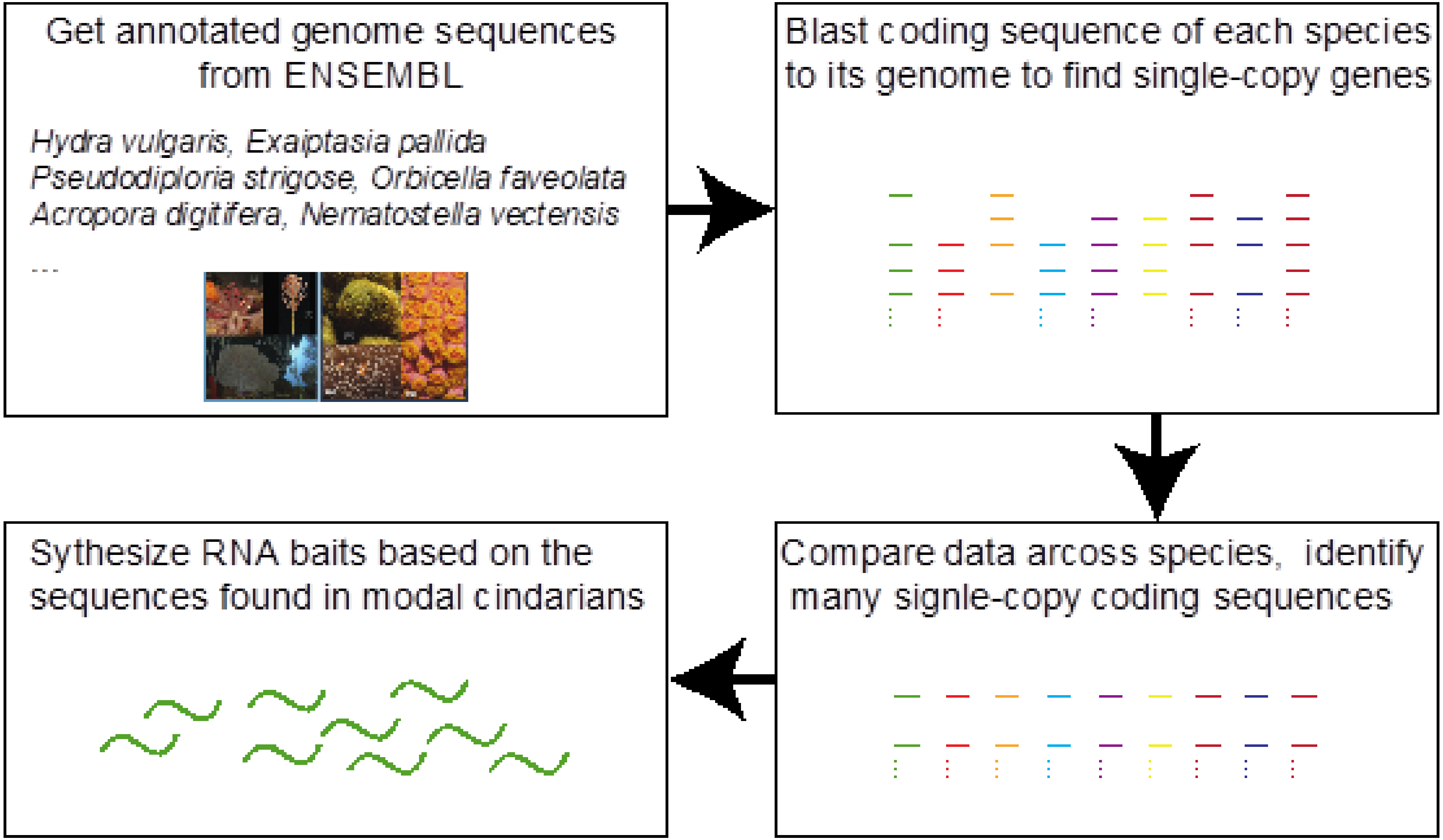
et al., 2010. Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. BMC Evolutionary Biology, 10:90.
et al., 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evolutionary Biology, 44.

you want to read more about how to search the database, please see Document, Demo, and FAQs. A Taxa List summarizes the species that are vailable for searching in the current version of EvolMarkers. If you prefer to build your own database, the scripts can be downloaded a EvolMarkers ools.

you are ready to quest your silver bullets, just go **Searching Markers**.

or questions, please contact Chenhong Li (lichenhong.unl@gmail.com).
ack to the Main Page.

A BIGGER PICTURE



Get annotated genome sequences from ENSEMBL

Hydra vulgaris, Exaiptasia pallida
Pseudodiploria strigose, Orbicella faveolata
Acropora digitifera, Nematostella vectensis
...

Blast coding sequence of each species to its genome to find single-copy genes

Compare data arcoss species, identify many signle-copy coding sequences

Sythesize RNA baits based on the sequences found in modal cindarians

Main Steps

- Constructed database
- Search database
- Filter
- Predict Frame

Constructed database

- Genome sequences
  - fna
  - genbank

- Annotation
  - embl
  - gff

e!Ensembl

http://asia.ensembl.org/index.html

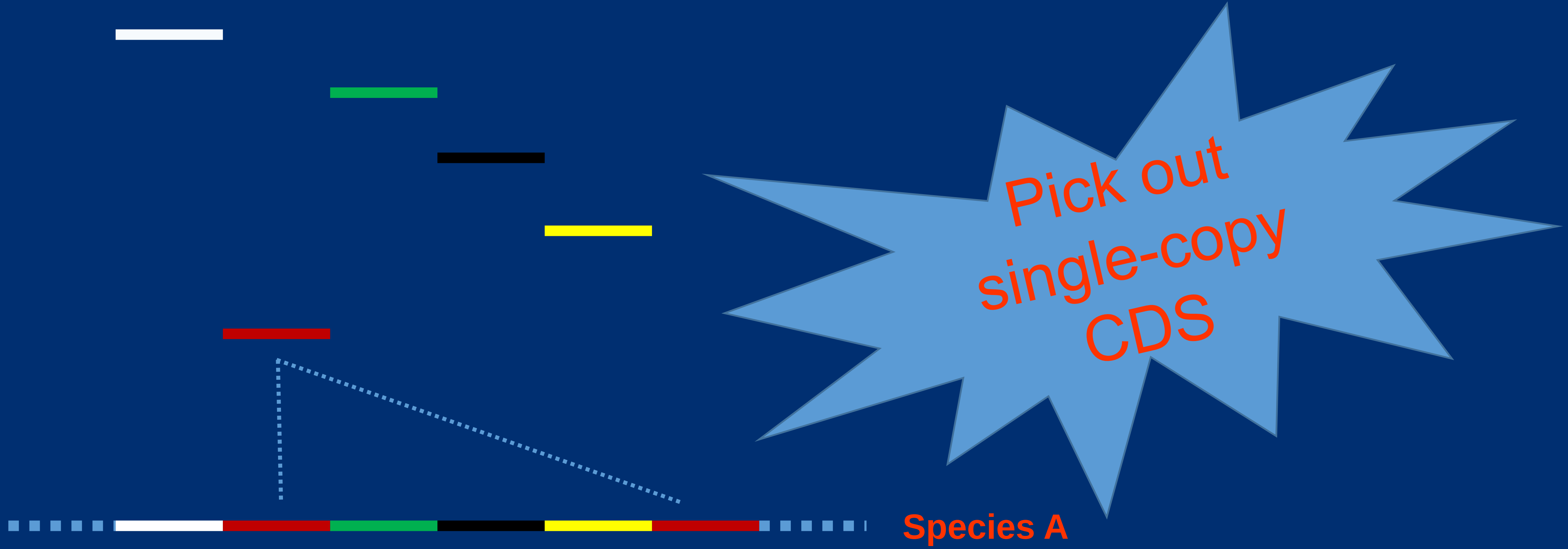
NCBI
National Center for
Biotechnology Information

https://www.ncbi.nlm.nih.gov/

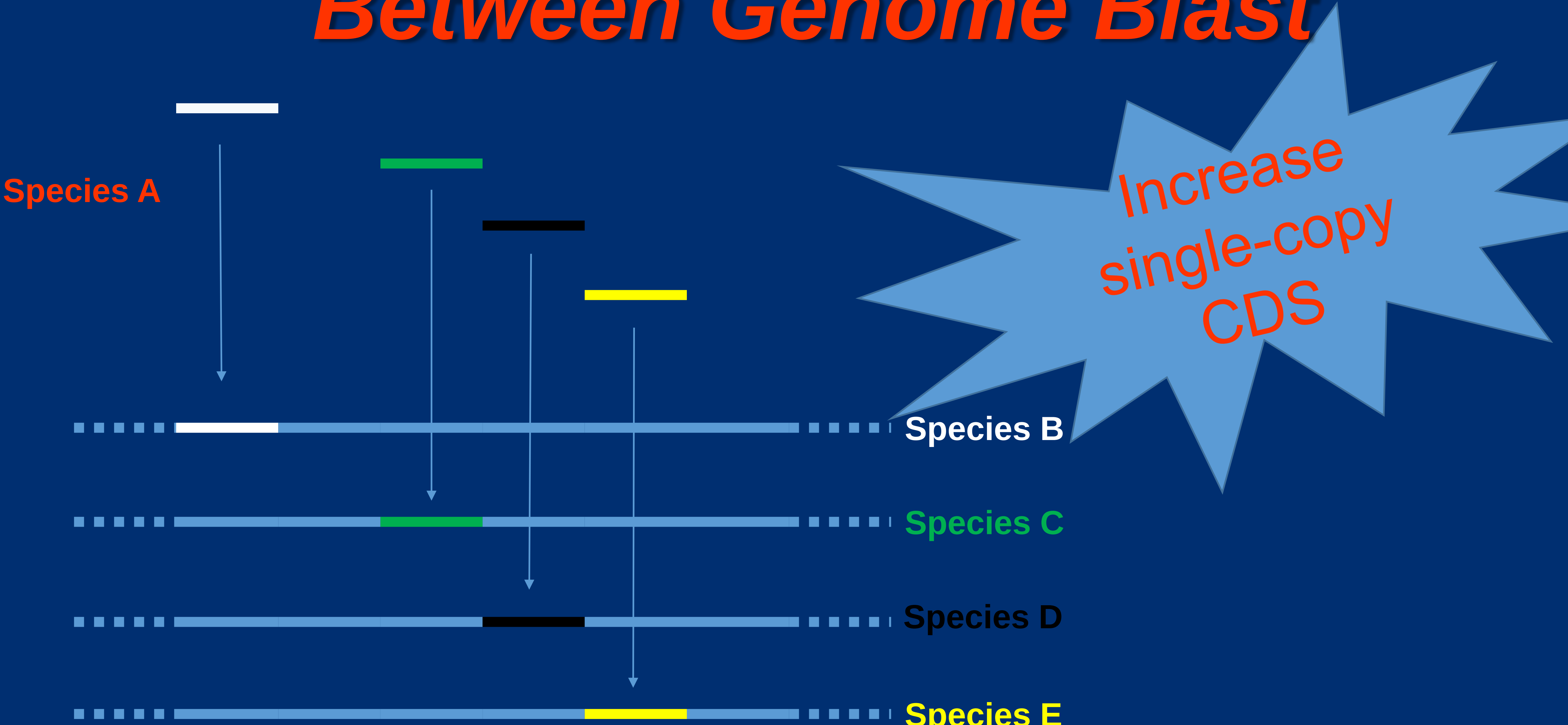
Within Genome Blast

Between Genome Blast

Within Genome Blast



Between Genome Blast



Filter by Programming & Manual

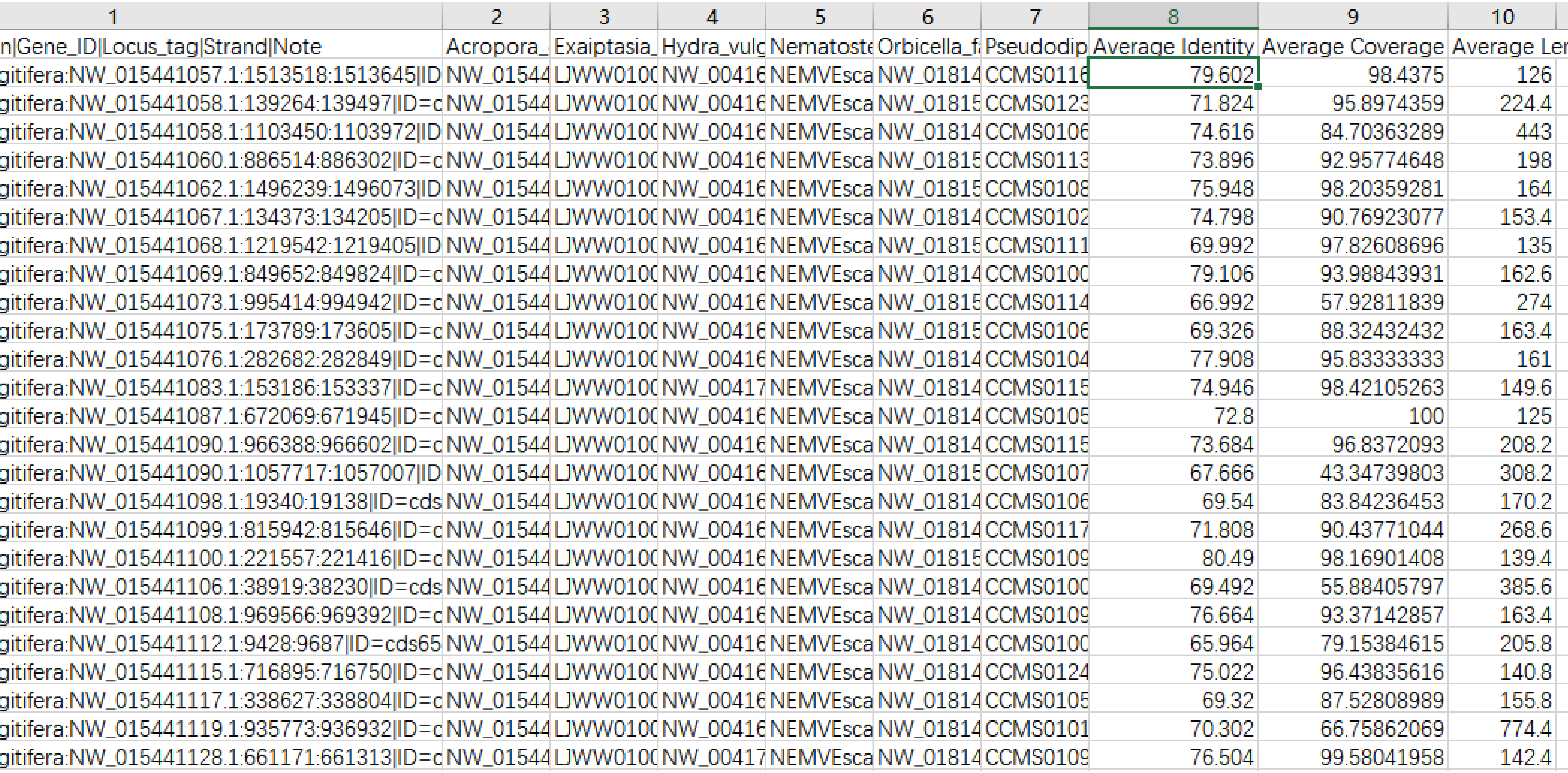
- Find markers
- Sort markers
- Parse markers

A.onehitCDSmarkers.txt
B.onehitCDSmarkers.txt
C.onehitCDSmarkers.txt
D.onehitCDSmarkers.txt

REDUNDANT

refined

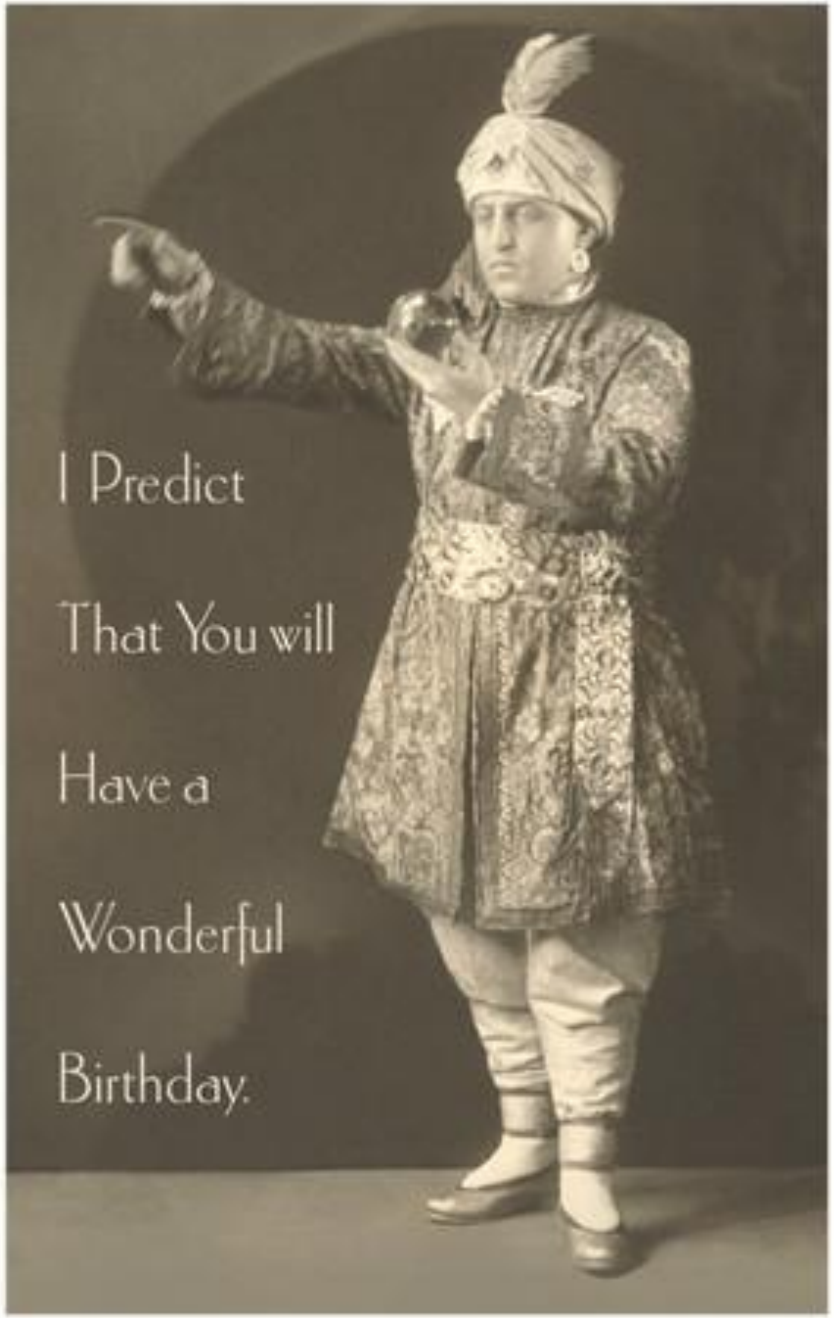
A.CDSmarkers.fas
B.CDSmarkers.fas
C.CDSmarkers.fas
D.CDSmarkers.fas

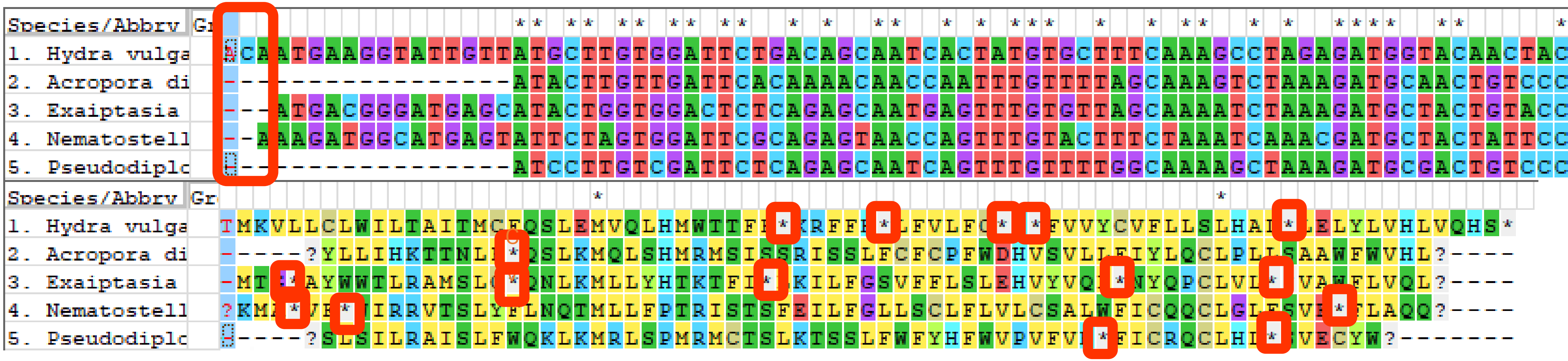


1	2	3	4	5	6	7	8	9	10
n|Gene_ID|Locus_tag|Strand|Note	Acropora_	Exaiptasia_	Hydra_vulg	Nematoste	Orbicella_fa	Pseudodip	Average Identity	Average Coverage	Average Ler
gitifera:NW_015441057.1:1513518:1513645|ID	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0116	79.602	98.4375	126
gitifera:NW_015441058.1:139264:139497|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01815	CCMS0123	71.824	95.8974359	224.4
gitifera:NW_015441058.1:1103450:1103972|ID	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0106	74.616	84.70363289	443
gitifera:NW_015441060.1:886514:886302|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01815	CCMS0113	73.896	92.95774648	198
gitifera:NW_015441062.1:1496239:1496073|ID	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01815	CCMS0108	75.948	98.20359281	164
gitifera:NW_015441067.1:134373:134205|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0102	74.798	90.76923077	153.4
gitifera:NW_015441068.1:1219542:1219405|ID	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01815	CCMS0111	69.992	97.82608696	135
gitifera:NW_015441069.1:849652:849824|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0100	79.106	93.98843931	162.6
gitifera:NW_015441073.1:995414:994942|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01815	CCMS0114	66.992	57.92811839	274
gitifera:NW_015441075.1:173789:173605|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01815	CCMS0106	69.326	88.32432432	163.4
gitifera:NW_015441076.1:282682:282849|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0104	77.908	95.83333333	161
gitifera:NW_015441083.1:153186:153337|ID=c	NW_01544	LJWW0100	NW_00417	NEMVEsca	NW_01814	CCMS0115	74.946	98.42105263	149.6
gitifera:NW_015441087.1:672069:671945|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0105	72.8	100	125
gitifera:NW_015441090.1:966388:966602|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0115	73.684	96.8372093	208.2
gitifera:NW_015441090.1:1057717:1057007|ID	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01815	CCMS0107	67.666	43.34739803	308.2
gitifera:NW_015441098.1:19340:19138|ID=cds	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0106	69.54	83.84236453	170.2
gitifera:NW_015441099.1:815942:815646|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0117	71.808	90.43771044	268.6
gitifera:NW_015441100.1:221557:221416|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01815	CCMS0109	80.49	98.16901408	139.4
gitifera:NW_015441106.1:38919:38230|ID=cds	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0100	69.492	55.88405797	385.6
gitifera:NW_015441108.1:969566:969392|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0109	76.664	93.37142857	163.4
gitifera:NW_015441112.1:9428:9687|ID=cds65	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0100	65.964	79.15384615	205.8
gitifera:NW_015441115.1:716895:716750|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0124	75.022	96.43835616	140.8
gitifera:NW_015441117.1:338627:338804|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0105	69.32	87.52808989	155.8
gitifera:NW_015441119.1:935773:936932|ID=c	NW_01544	LJWW0100	NW_00416	NEMVEsca	NW_01814	CCMS0101	70.302	66.75862069	774.4
gitifera:NW_015441128.1:661171:661313|ID=c	NW_01544	LJWW0100	NW_00417	NEMVEsca	NW_01814	CCMS0109	76.504	99.58041958	142.4

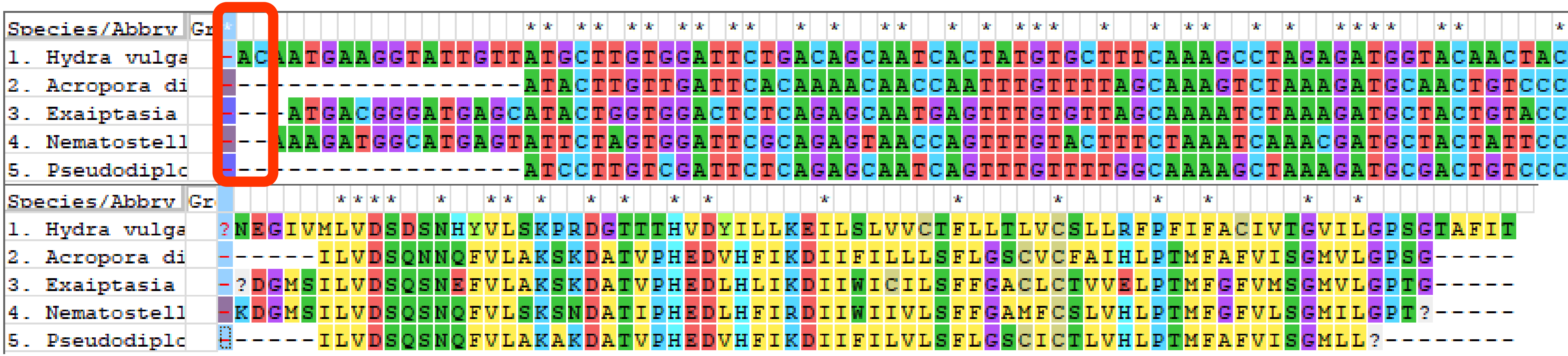
Predict Frame

- Get a full-coding sequence
- Useful for blastp





```
Species/Abbrv Gr                    ** ** ** ** **  * *  **  * * ***  *  * **  * *  ****  **     *
1. Hydra vulga    ACAATGAAGGTATTGTTATGCTTGTGGATTCTGACAGCAATCACTATGTGCTTTCAAAGCCTAGAGATGGTACAACTAC
2. Acropora di    -----------------ATACTTGTTGATTCACAAAACAACCAATTTGTTTTAGCAAAGTCTAAAGATGCAACTGTCCC
3. Exaiptasia     ---ATGACGGGATGAGCATACTGGTGGACTCTCAGAGCAATGAGTTTGTGTTAGCAAAATCTAAAGATGCTACTGTACC
4. Nematostell    --AAAGATGGCATGAGTATTCTAGTGGATTCGCAGAGTAACCAGTTTGTACTTTCTAAATCAAACGATGCTACTATTCC
5. Pseudodiplc    -----------------ATCCTTGTCGATTCTCAGAGCAATCAGTTTGTTTTGGCAAAAGCTAAAGATGCGACTGTCCC
Species/Abbrv Gr                        *                                    *
1. Hydra vulga    TMKVLLCLWILTAITMCFQSLEMVQLHMWTTFF*KRFFH*LFVLFC*[illegible]*FVVYCVFLLSLHAI*LELYLVHLVQHS*
2. Acropora di    -----?YLLIHKTTNLI*QSLKMQLSHMRMSISSRISSLFCFCPFWDHVSVLLFIYLQCLPLLSAAWFWVHL?----
3. Exaiptasia     -MT[illegible]*AYWWTLRAMSLC*QNLKMLLYHTKTFI*[illegible]KILFGSVFFLSLEHVYVQI*NYQPCLVL*[illegible]VAWFLVQL?----
4. Nematostell    ?KMA*VF*[illegible]IRRVTSLYFLNQTMLLFPTRISTSFEILFGLLSCLFLVLCSALWFICQQCLGLFSVF*FLAQQ?----
5. Pseudodiplc    -----?SLSILRAISLFWQKLKMRLSPMRMCTSLKTSSLFWFYHFWVPVFVI*FICRQCLHI*SVECYW?-------
```



```
Species/Abbrv Gr  *                  ** ** ** ** **  * *  **  * * ***  *  * **  * *  ****  **     *
1. Hydra vulga    -ACAATGAAGGTATTGTTATGCTTGTGGATTCTGACAGCAATCACTATGTGCTTTCAAAGCCTAGAGATGGTACAACTAC
2. Acropora di    ------------------ATACTTGTTGATTCACAAAACAACCAATTTGTTTTAGCAAAGTCTAAAGATGCAACTGTCCC
3. Exaiptasia     ----ATGACGGGATGAGCATACTGGTGGACTCTCAGAGCAATGAGTTTGTGTTAGCAAAATCTAAAGATGCTACTGTACC
4. Nematostell    ---AAAGATGGCATGAGTATTCTAGTGGATTCGCAGAGTAACCAGTTTGTACTTTCTAAATCAAACGATGCTACTATTCC
5. Pseudodiplc    ------------------ATCCTTGTCGATTCTCAGAGCAATCAGTTTGTTTTGGCAAAAGCTAAAGATGCGACTGTCCC
Species/Abbrv Gr         ****  *  ** *  * *  * *      *       *     *     *  *     *  *
1. Hydra vulga    ?NEGIVMLVDSDSNHYVLSKPRDGTTTHVDYILLKEILSLVVCTFLLTLVCSLLRFPFIFACIVTGVILGPSGTAFIT
2. Acropora di    ------ILVDSQNNQFVLAKSKDATVPHEDVHFIKDIIFILLLSFLGSCVCFAIHLPTMFAFVISGMVLGPSG-----
3. Exaiptasia     -?DGMSILVDSQSNEFVLAKSKDATVPHEDLHLIKDIIWICILSFFGACLCTVVELPTMFGFVMSGMVLGPTG-----
4. Nematostell    -KDGMSILVDSQSNQFVLSKSNDATIPHEDLHFIRDIIWIIVLSFFGAMFCSLVHLPTMFGFVLSGMILGPT?-----
5. Pseudodiplc    ------ILVDSQSNQFVLAKAKDATVPHEDVHFIKDIIFILVLSFLGSCICTLVHLPTMFAFVISGMLL?--------
```

MYcroarray

MYbaits® Bait Design

Procedure and Data

A QUICK RECAP

“Constructed Database”

“Search Database”

“Filter”

“Predict Frame”


Thank You!