Genomic phylostratigraphy reveals a wealth of *Hermetia illucens* species-specific genes of unknown function

Momir Futo

Shcool of Medicine
Catholic University of Croatia

Athens, 29.1.2025.

Genomic phylostratigraphy

- genetic bio-statistical method
- dating the origin of specific genes by looking at their homologs across species
- system links genes to their **founder gene**, allowing us to then determine their age
- better understand many evolutionary processes such as
  - patterns of gene birth throughout evolution
  - the relationship between the age of a transcriptome throughout embryonic development

Founder genes

- diversity of the genome is not caused only by gene duplications but also by continuous, frequent *de novo* gene births
- "founder genes" - form from non-genic DNA sequences, as well as from changes in reading frame, other ways of arising from existing genes
- the founder genes can then be sorted into a specific **phylostratum** - the clade that includes all the genes that derive from the same founder gene
- we can signify that this gene was formed in the common ancestor of a certain clade
- (e.g. Arthropoda, Mammalia, Metazoa, etc.)

Evolutionary age and function

- Positioning the founder genes and their descendants on different phylostrata allows us to age them
- This can then be used to analyze the origin of certain functions of proteins and developmental processes on a macroevolutionary scale

A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns

Tomislav Domazet-Lošo ✉ & Diethard Tautz

Nature **468**, 815–818 (2010) | Cite this article

> Trends Genet. 2007 Nov;23(11):533-9. doi: 10.1016/j.tig.2007.08.014.

A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages

Tomislav Domazet-Loso [1], Josip Brajković, Diethard Tautz

Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis

Naoki Irie ✉ & Shigeru Kuratani

Nature Communications **2**, Article number: 248 (2011) | Cite this article

A transcriptomic hourglass in plant embryogenesis

Marcel Quint ✉, Hajk-Georg Drost, Alexander Gabel, Kristian Karsten Ullrich, Markus Bönn & Ivo Grosse

Nature **490**, 98–101 (2012) | Cite this article

Mol Biol Evol. 2021 Jan 4;38(1):31-47. doi: 10.1093/molbev/msaa217.

Embryo-Like Features in Developing Bacillus subtilis Biofilms

Momir Futo[1], Luka Opašić[1,2], Sara Koska[1], Nina Čorak[1], Tin Široki[3], Vaishnavi Ravikumar[4], Annika Thorsell[5], Maša Lenuzzi[1,6], Domagoj Kifer[7], Mirjana Domazet-Lošo[3], Kristian Vlahoviček[8,9], Ivan Mijakovic[4,10], Tomislav Domazet-Lošo[1,11]

Mol Biol Evol. 2020 Jun 1;37(6):1667-1678. doi: 10.1093/molbev/msaa035.

Evolutionary Analysis of the Bacillus subtilis Genome Reveals New Genes Involved in Sporulation

Lei Shi[1], Abderahmane Derouiche[1], Santosh Pandit[1], Shadi Rahimi[1], Aida Kalantari[1], Momir Futo[2], Vaishnavi Ravikumar[3], Carsten Jers[3], Venkata R S S Mokkapati[1], Kristian Vlahoviček[4,5], Ivan Mijakovic[1,3]

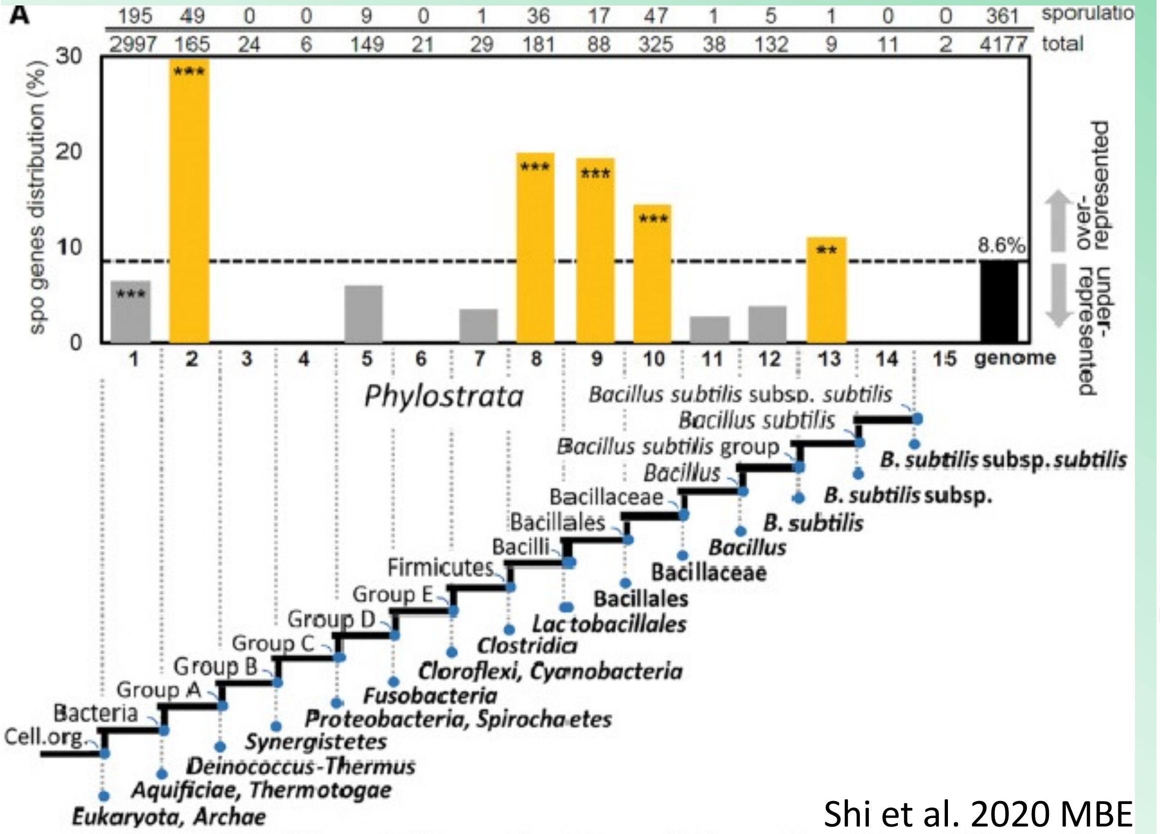
Article

Pleomorphic Variants of *Borreliella* (syn. *Borrelia*) *burgdorferi* Express Evolutionary Distinct Transcriptomes

Nina Čorak [1], Sirli Anniko [2,3], Christina Daschkin-Steinborn [2], Viktoria Krey [2,4], Sara Koska [1], Momir Futo [1,5,6], Tin Široki [5], Innokenty Woichansky [2], Luka Opašić [1], Domagoj Kifer [7], Anja Tušar [1], Horst-Günter Maxeiner [2,8], Mirjana Domazet-Lošo [5], Carsten Nicolaus [2] and Tomislav Domazet-Lošo [1,6,*]

Phylostratigraphic map

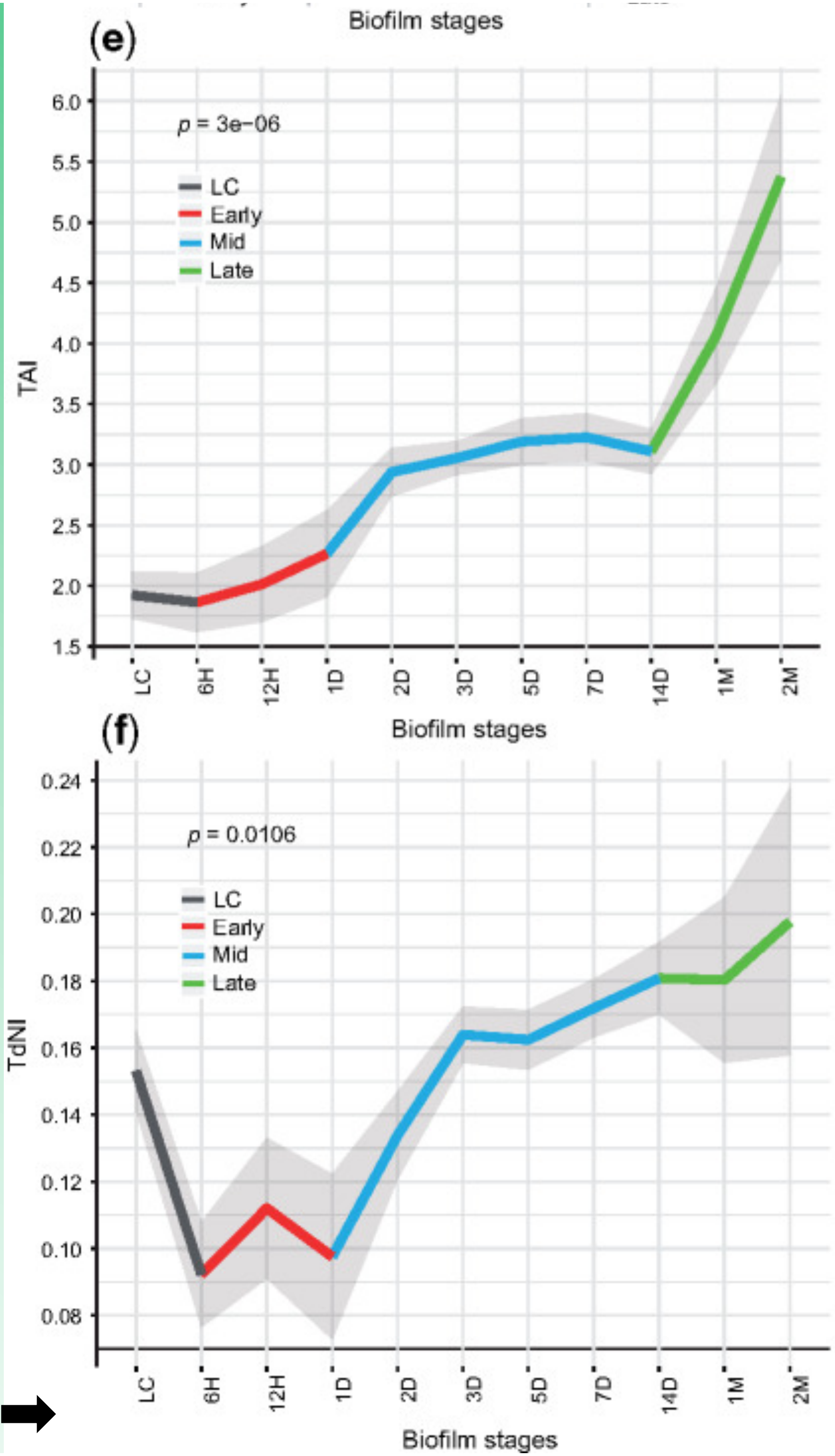
Predictive power ⬇



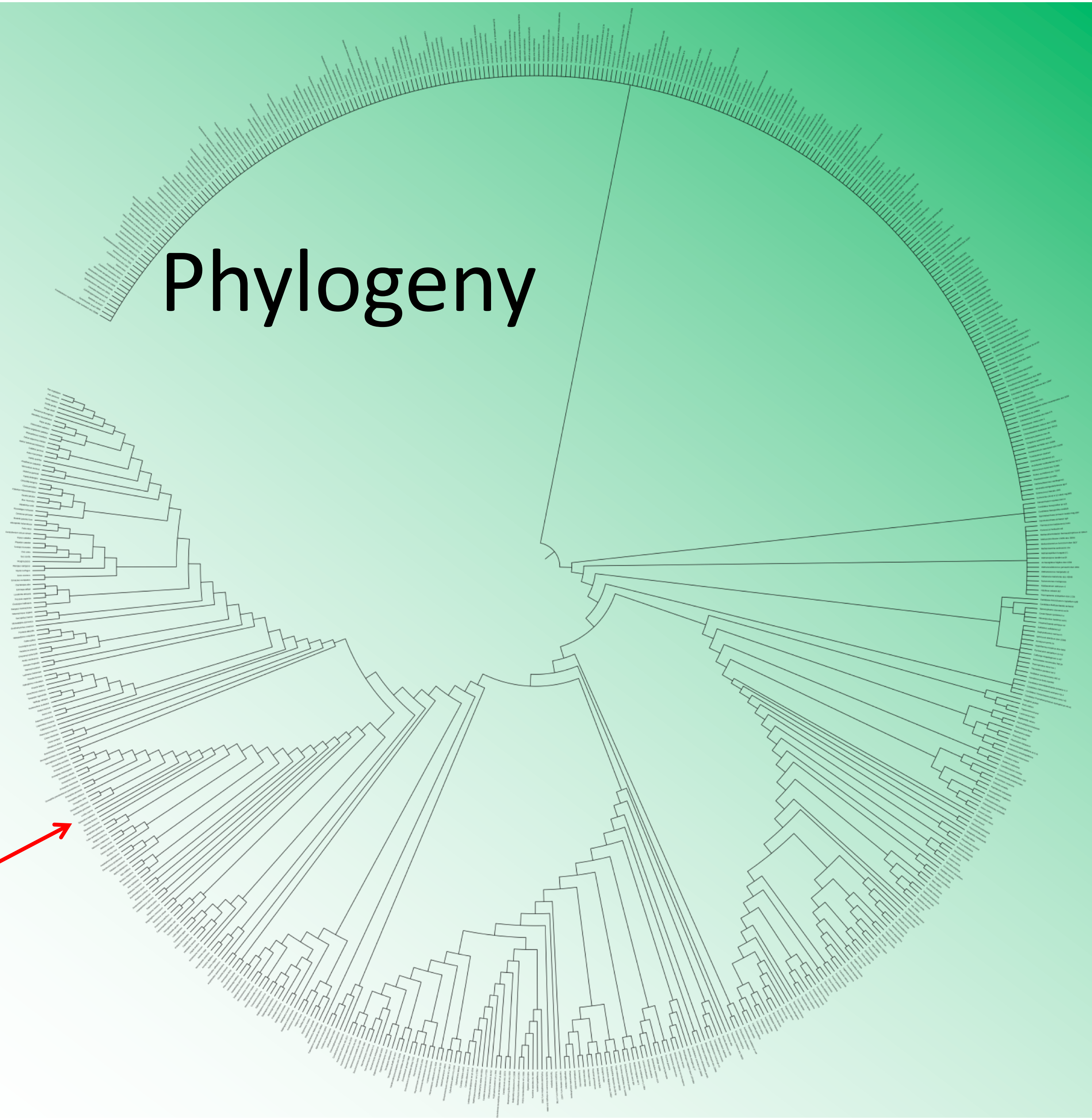
Shi et al. 2020 MBE

Futo et al. 2021 MBE

Developmental transcriptomics ➡



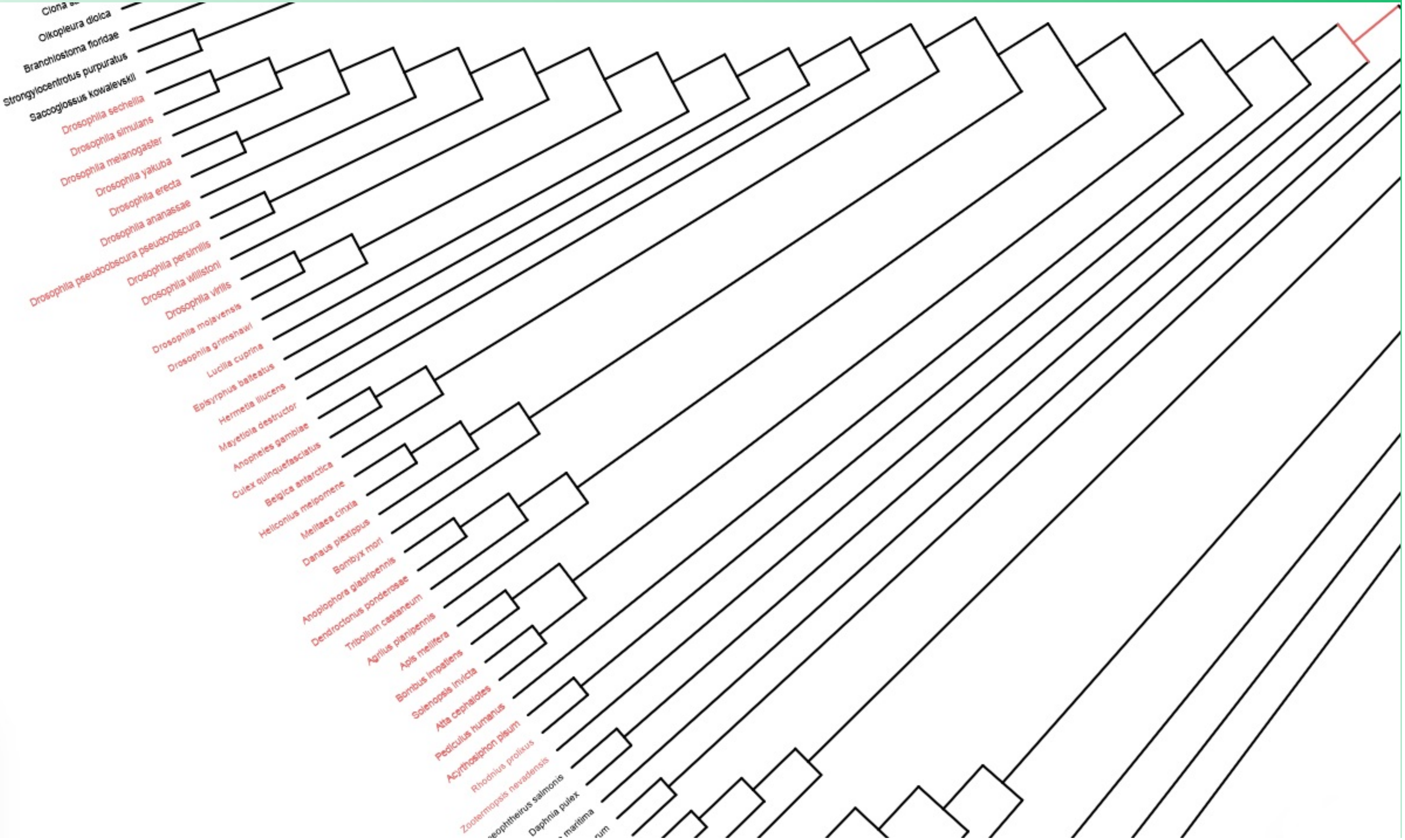
Phylogeny



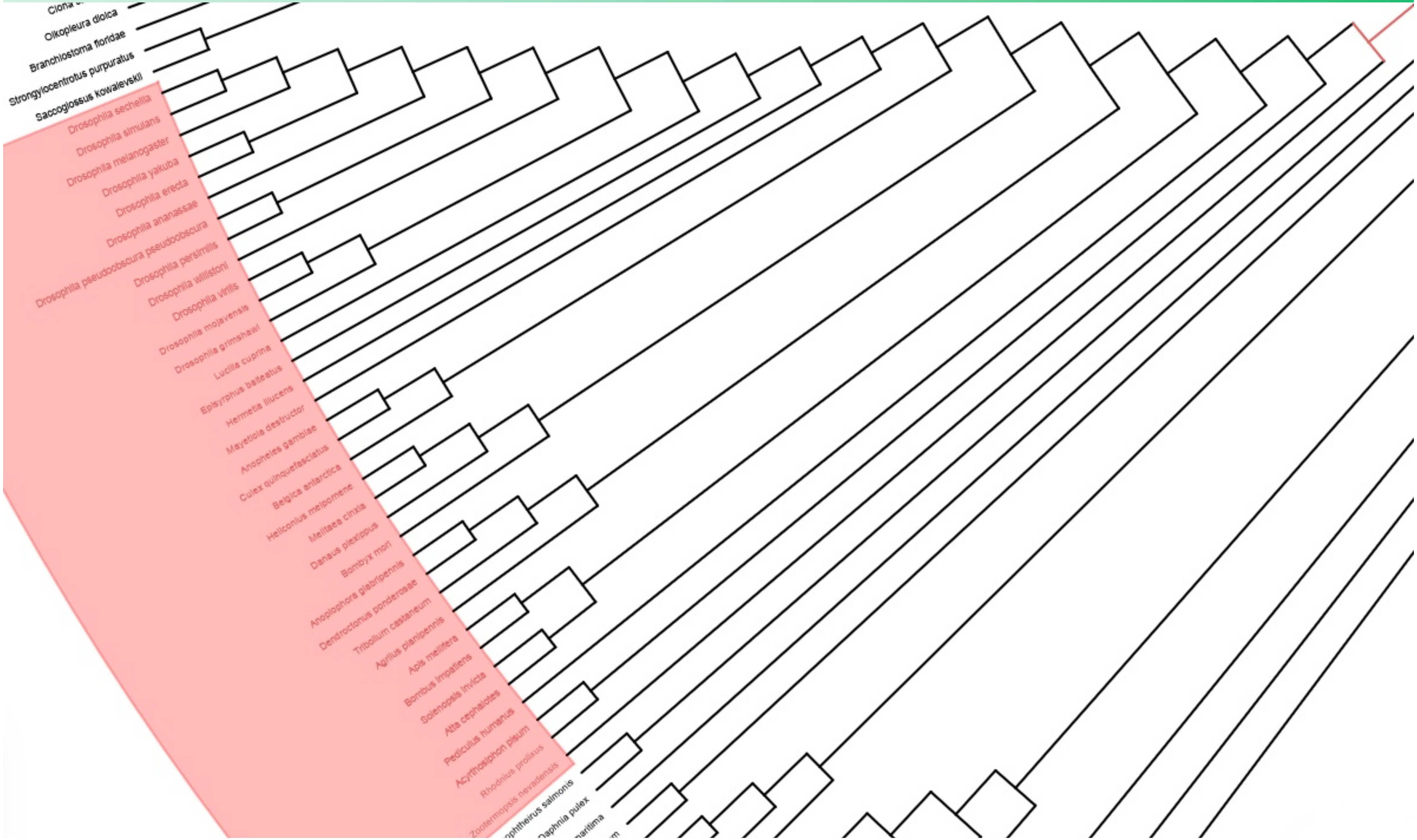
BSF

668 species

Insecta – 35 species



Insecta – 35 species



Our phylostratigraphic map

Phylostratum	**PS_No.**	**No. of proteins**
Cellular_organisms	1	4839
DPANN/Euryarchaeota/TACK/Asgard_archaea/Eukaryota	2	123
Euryarchaeota/TACK/Asgard_archaea/Eukaryota	3	50
TACK/Asgard_archaea/Eukaryota	4	57
Asgard_archaea/Eukaryota	5	69
Eukaryota	6	3901
Amorpheae	7	96
Obazoa	8	30
Opisthokonta	9	282
Holozoa	10	151
Filozoa	11	61
Choanoflagellida/Metazoa	12	267
Metazoa	13	295
Myriazoa	14	202
Parahoxozoa	15	357
Bilateria	16	441
Protostomia	17	126
Ecdysozoa	18	11
Arthropoda	19	183
Mandibulata	20	115
Pancrustacea	21	153
Insecta	22	438
Eumetabola	23	117
Holometabola/Psocodea	24	36
Holometabola	25	92
Aparaglossata	26	160
Mecopterida	27	44
Diptera	28	295
Bibionomorpha	29	42
Brachycera	30	213
Hermetia_illucens	31	**835**

H. illucens proteome:
14,081 protein entries

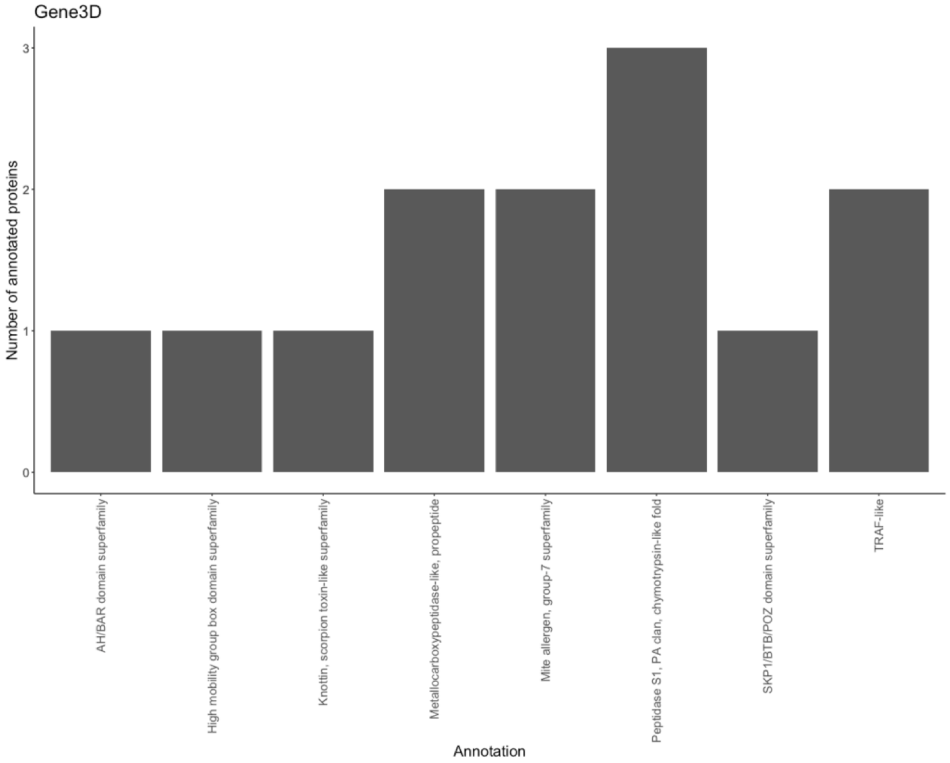
31 phylostrata in total

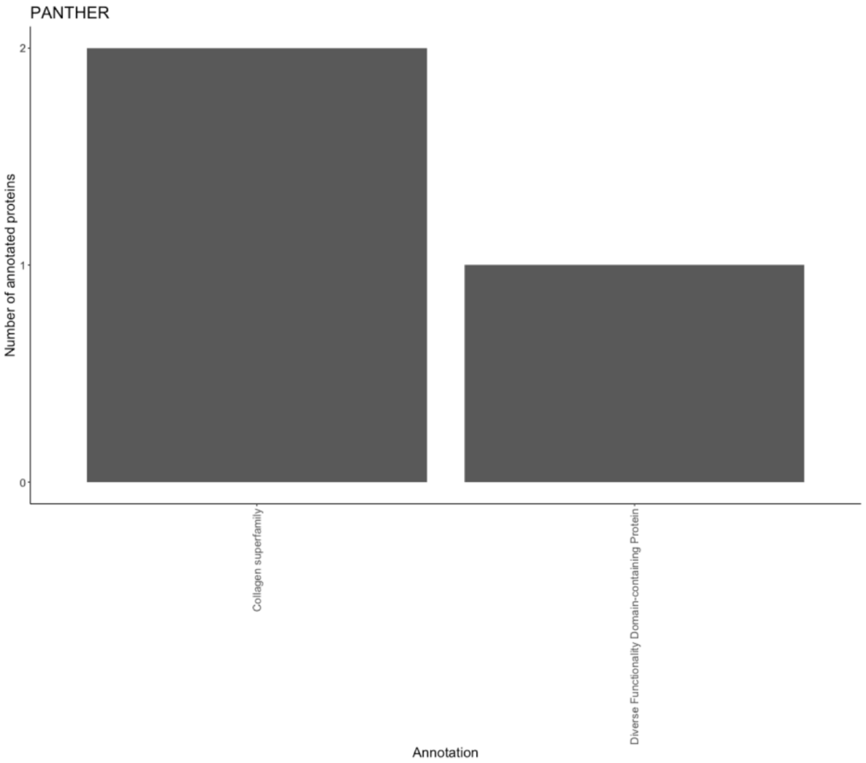
PS No. 31 – proteins specific to ***H. illucens***

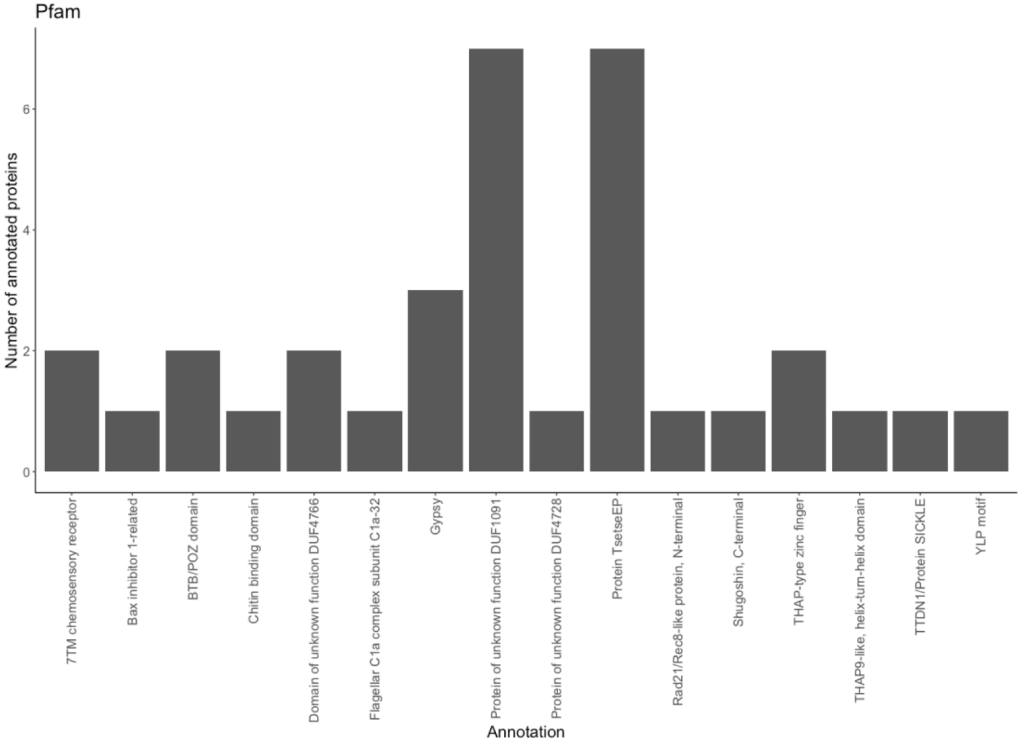
835

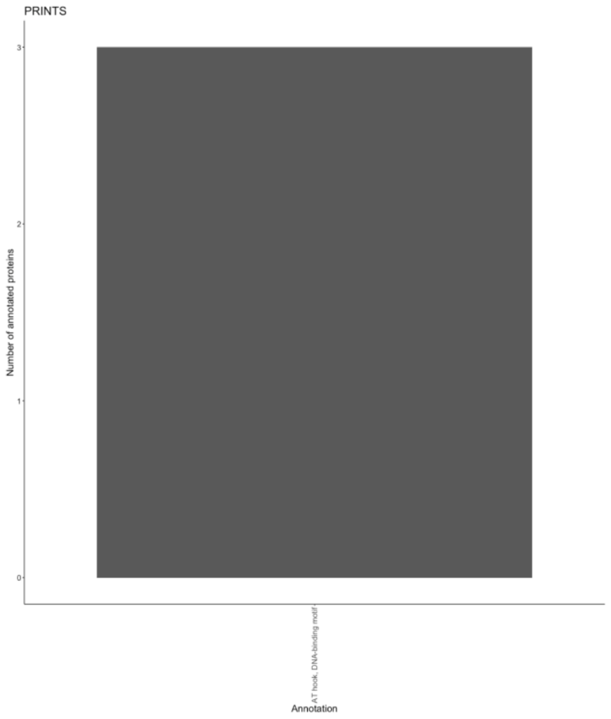
Functional analasys

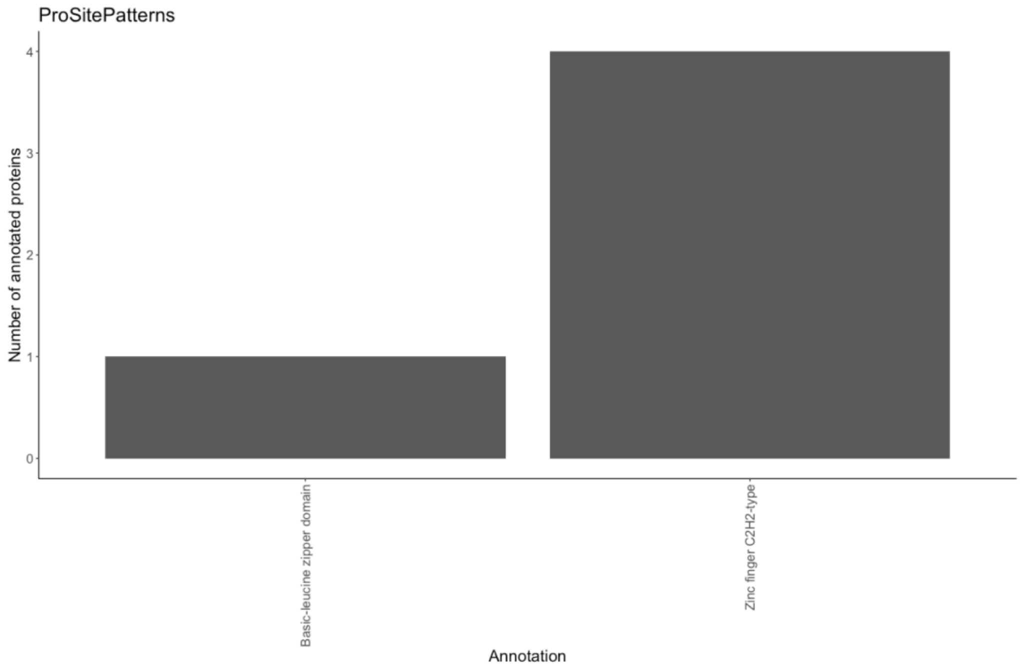
- 835 genes analyzed
- **EggNOG database** – 0 annotations
- **INterPro databases** – only a few anotations
    - **Gene3D**
    - **PANTHER**
    - **PRINTS**
    - **Pfam**
    - **ProSitePatterns**
    - **ProSiteProfiles**
    - **SMART**
    - **SUPERFAMILY**

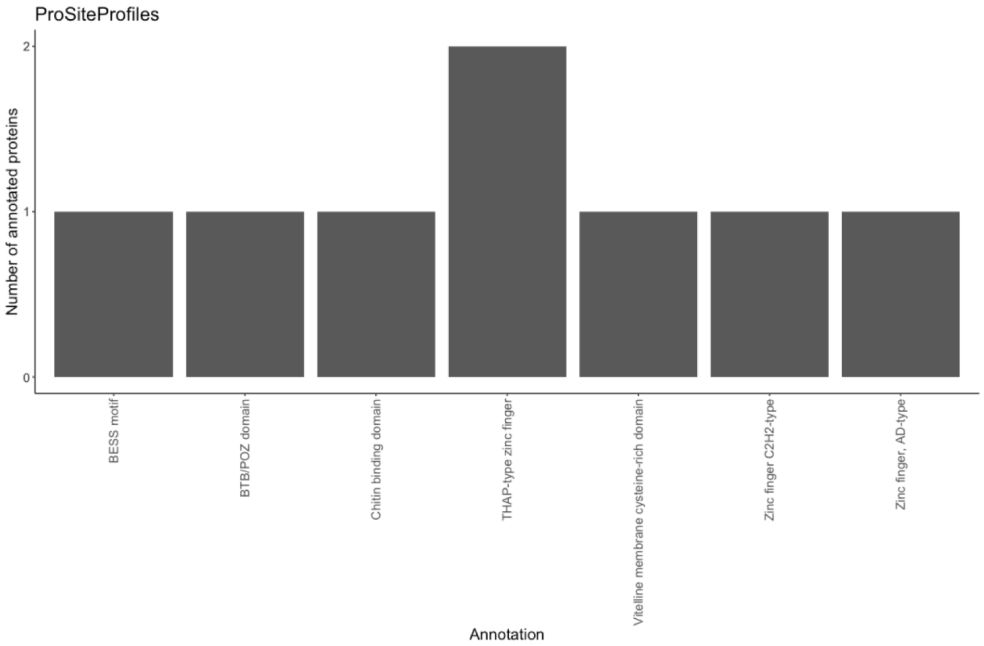

Gene3D
Number of annotated proteins
0
1
2
3
AH/BAR domain superfamily
High mobility group box domain superfamily
Knottin, scorpion toxin-like superfamily
Metallocarboxypeptidase-like, propeptide
Mite allergen, group-7 superfamily
Peptidase S1, PA clan, chymotrypsin-like fold
SKP1/BTB/POZ domain superfamily
TRAF-like
Annotation

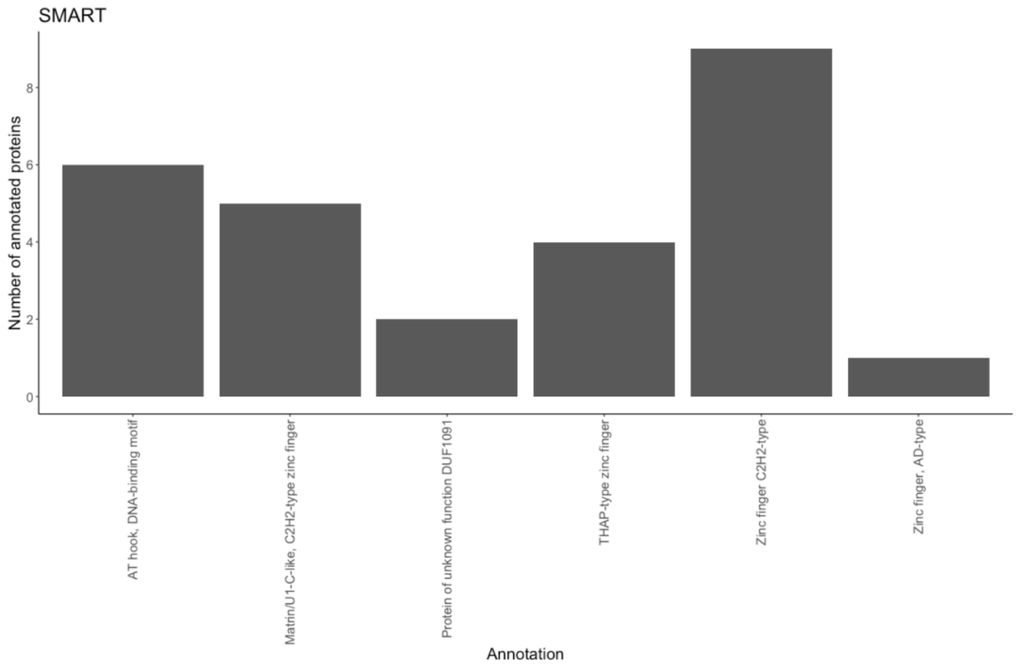

PANTHER
Number of annotated proteins
0
1
2
Collagen superfamily
Diverse Functionality Domain-containing Protein
Annotation

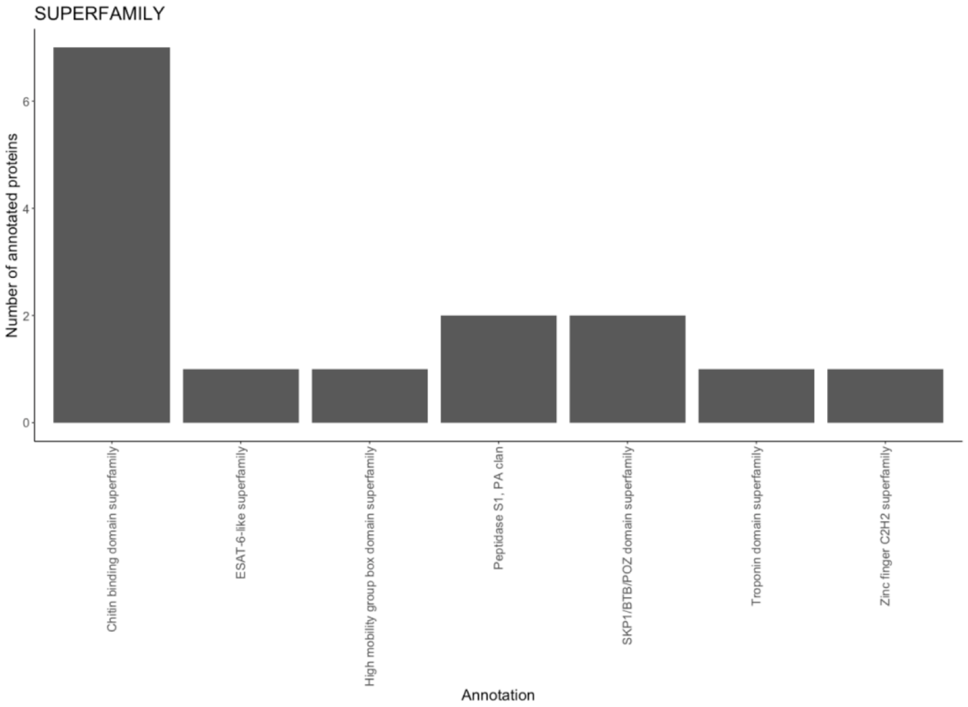

Pfam
Number of annotated proteins
0
2
4
6
7TM chemosensory receptor
Bax inhibitor 1-related
BTB/POZ domain
Chitin binding domain
Domain of unknown function DUF4766
Flagellar C1a complex subunit C1a-32
Gypsy
Protein of unknown function DUF1091
Protein of unknown function DUF4728
Protein TsetseEP
Rad21/Rec8-like protein, N-terminal
Shugoshin, C-terminal
THAP-type zinc finger
THAP9-like, helix-turn-helix domain
TTDN1/Protein SICKLE
YLP motif
Annotation


PRINTS
Number of annotated proteins
0
1
2
3
AT hook, DNA-binding motif
Annotation


ProSitePatterns
Number of annotated proteins
0
1
2
3
4
Basic-leucine zipper domain
Zinc finger C2H2-type
Annotation


ProSiteProfiles
Number of annotated proteins
0
1
2
BESS motif
BTB/POZ domain
Chitin binding domain
THAP-type zinc finger
Vitelline membrane cysteine-rich domain
Zinc finger C2H2-type
Zinc finger, AD-type
Annotation


SMART
Number of annotated proteins
0
2
4
6
8
AT hook, DNA-binding motif
Matrin/U1-C-like, C2H2-type zinc finger
Protein of unknown function DUF1091
THAP-type zinc finger
Zinc finger C2H2-type
Zinc finger, AD-type
Annotation


SUPERFAMILY
Number of annotated proteins
0
2
4
6
Chitin binding domain superfamily
ESAT-6-like superfamily
High mobility group box domain superfamily
Peptidase S1, PA clan
SKP1/BTB/POZ domain superfamily
Troponin domain superfamily
Zinc finger C2H2 superfamily
Annotation

Outlook

- Plenty of genes of unknown function to work with
- To build a robust and resolved phylogeny enriched with insect clades
- *H. illucens* developmental transcriptomics

LUX VERA

HRVATSKO
KATOLIČKO
SVEUČILIŠTE
ZAGREB
UNIVERSITAS
STUDIORUM
CATHOLICA
CROATICA
ZAGRABIA

Thanks!

Institut
Ruđer
Bošković

Niko Kasalo

Tomislav Domazet-Lošo,
PhD