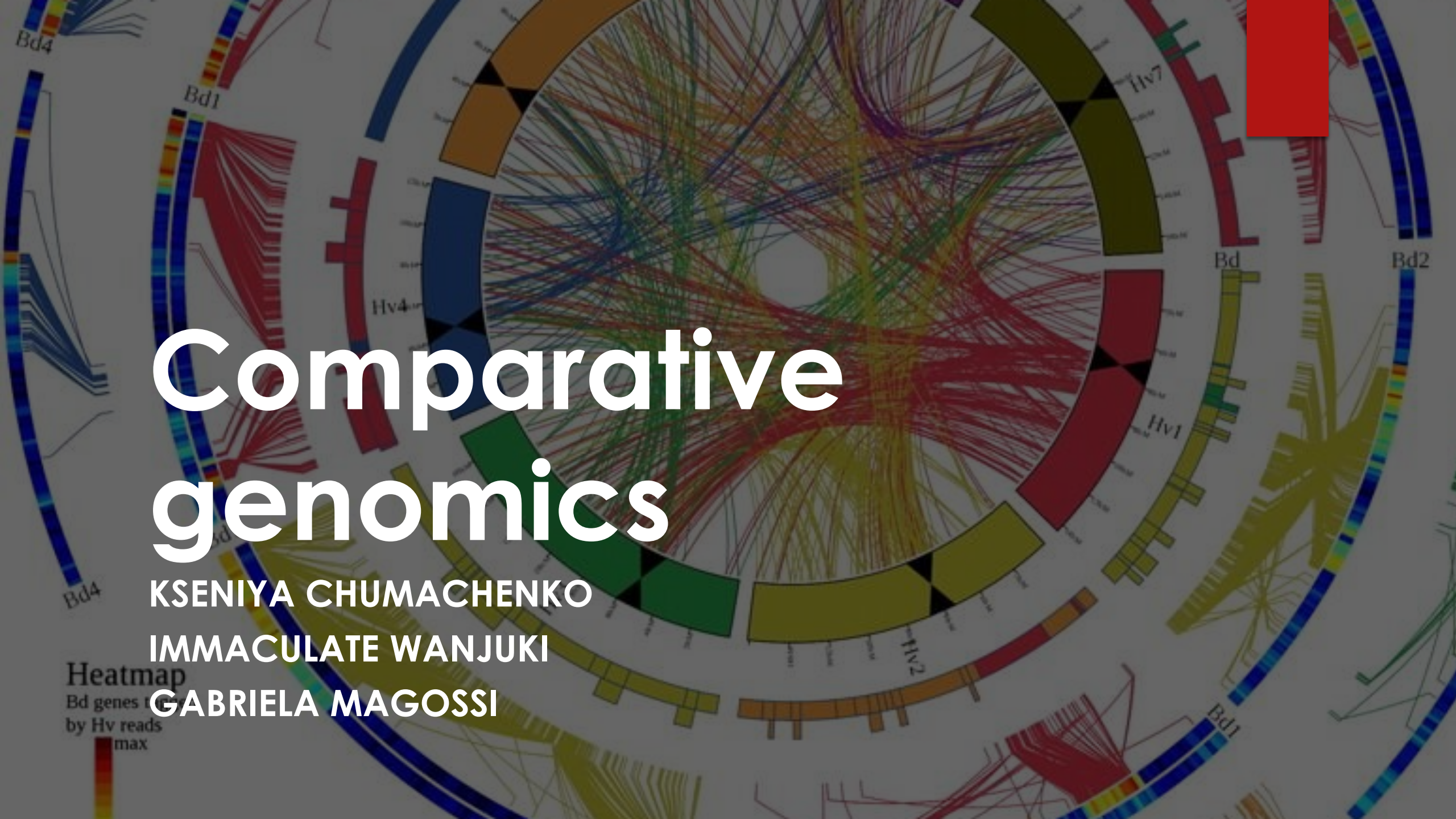


Comparative genomics

KSENIYA CHUMACHENKO

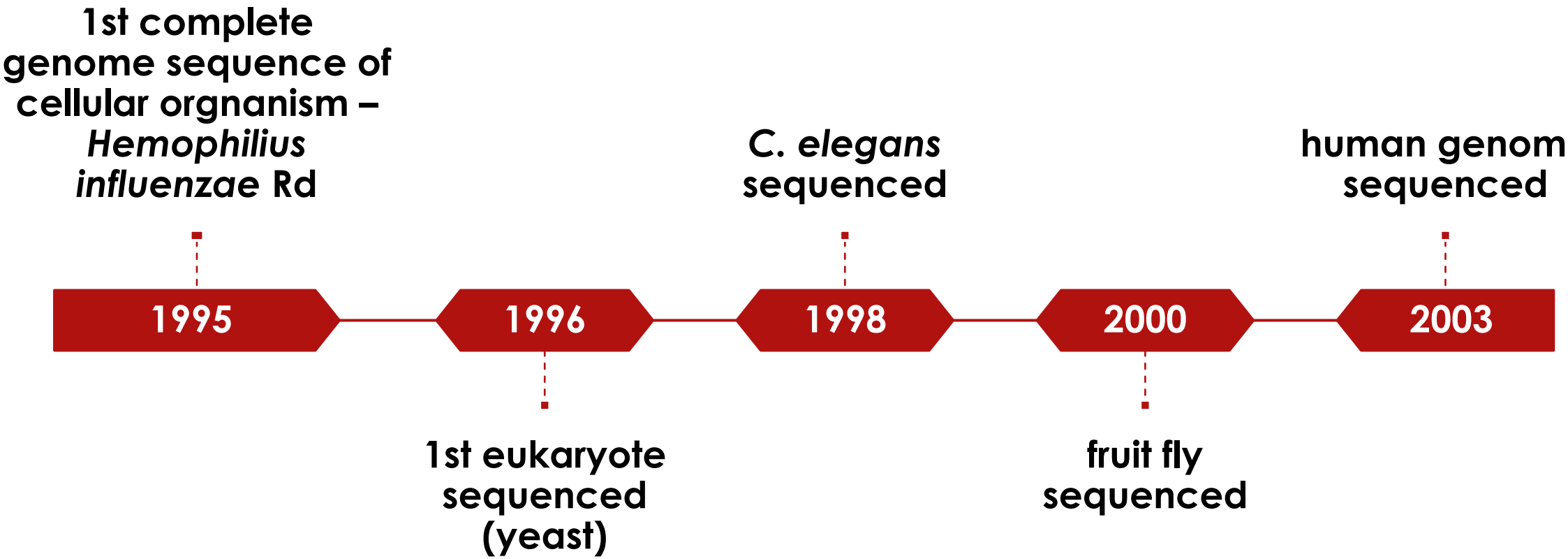
IMMACULATE WANJUKI

GABRIELA MAGOSSI

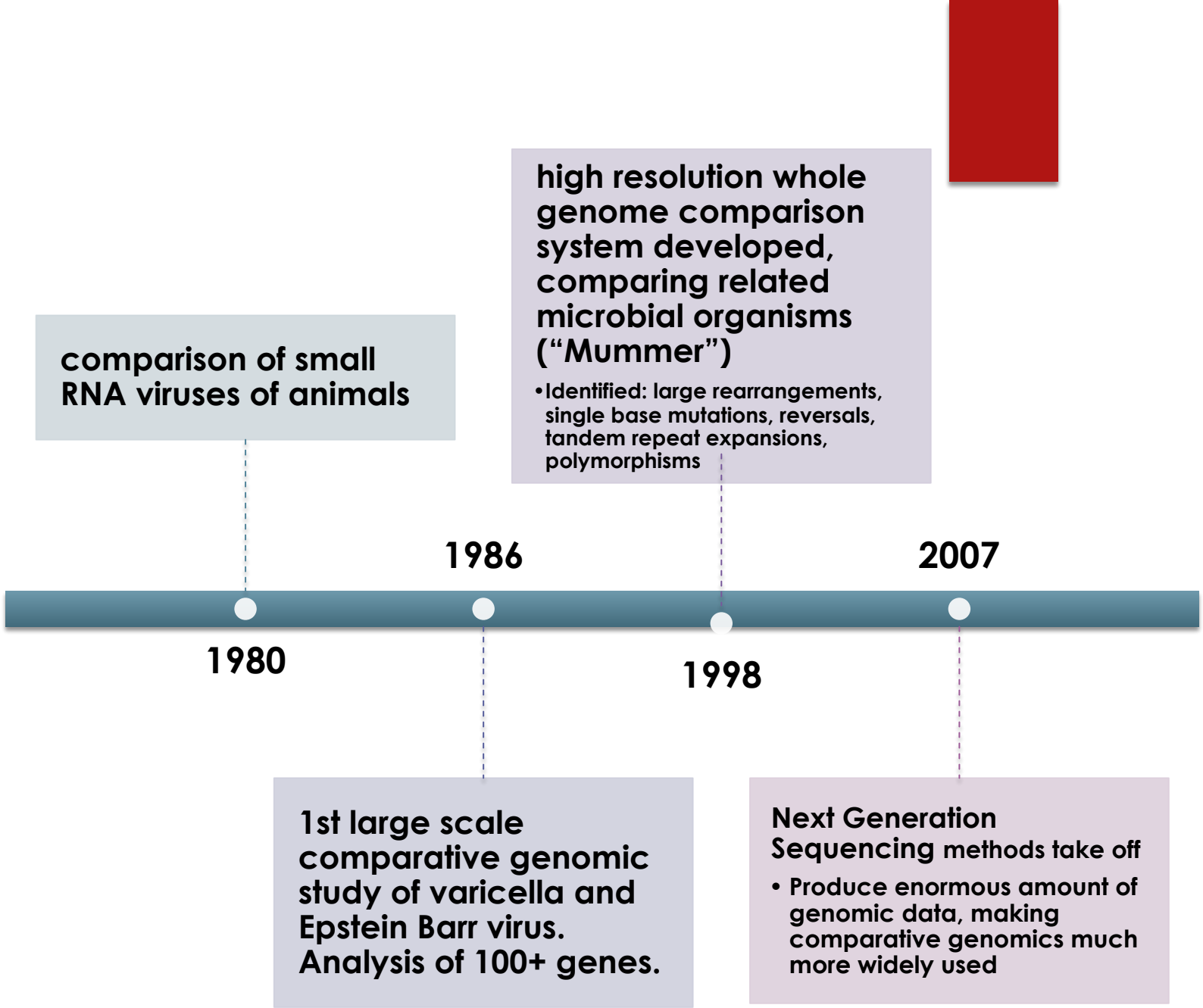
Presentation Outline

- Timeline/history
- Introduction
- Applications
- Workflow
  - Alignments
  - Comparative annotation
- Mini review of 1st comp. gen. paper
- Review #1
- Review #2
- Summary/Take aways

Timeline – 1st genomes sequenced



Brief history: Early Comparative Genomics



Early work setting the stage

"Comparative genomics of the eukaryotes"

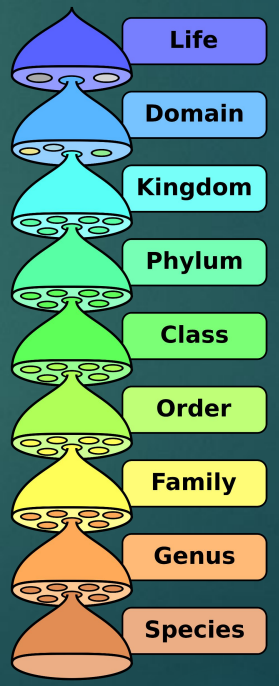
- *S. cerevisiae , C. elegans, D. melanogaster*

"Human and Mouse gene structure: comparative analysis and application to exon prediction"

Goals of Comparative genomics

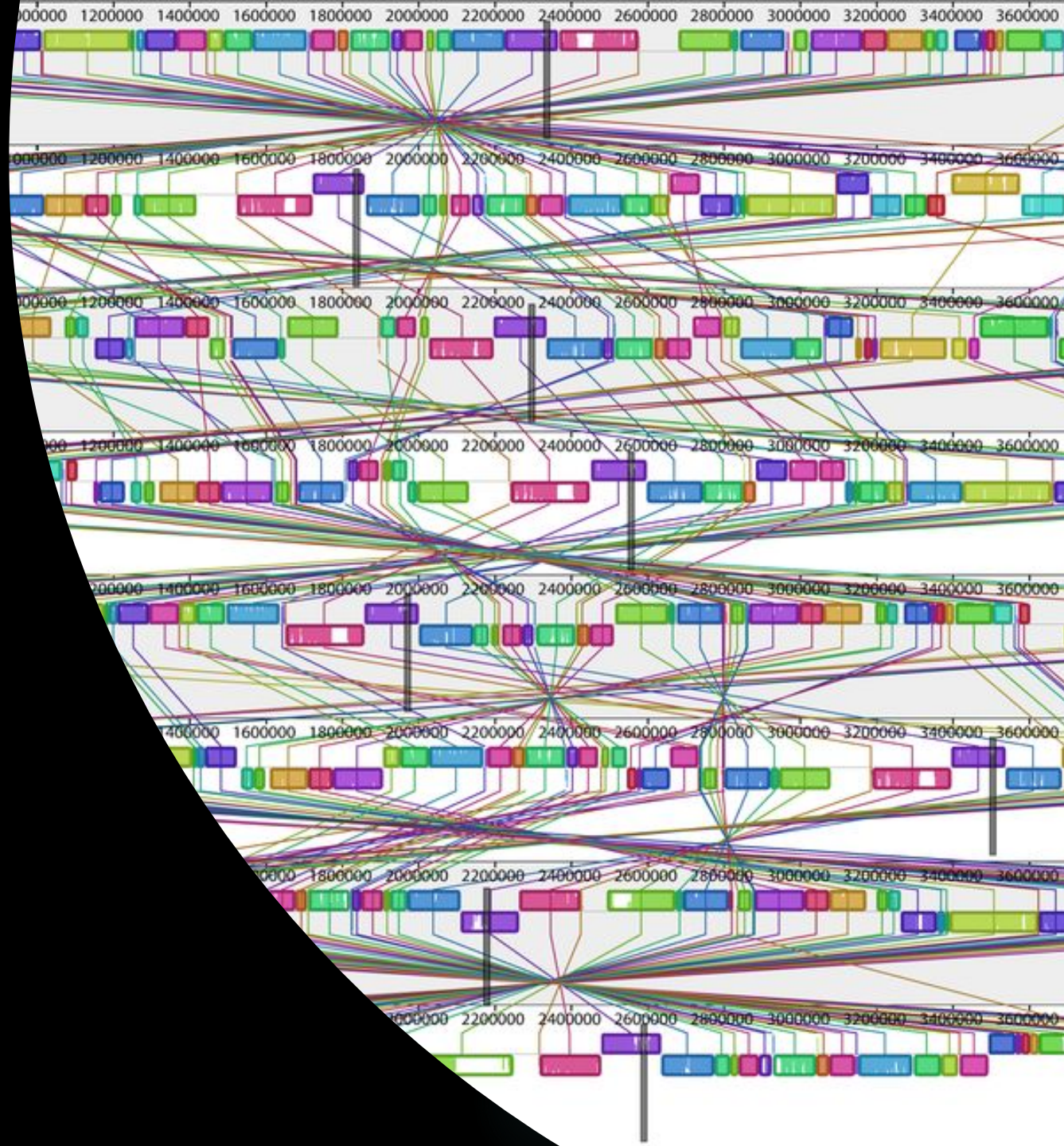
Comparison of **genomic features** of different **organisms**

- **Genomic features:**
  - DNA sequence
    - SNP and SV variation
  - Genes
  - Gene order (rearrangements)
  - Regulatory sequences
- **Between different organisms:**
  - Between different taxonomic groups
  - Within different taxonomic groups



Applications of comparative genomics

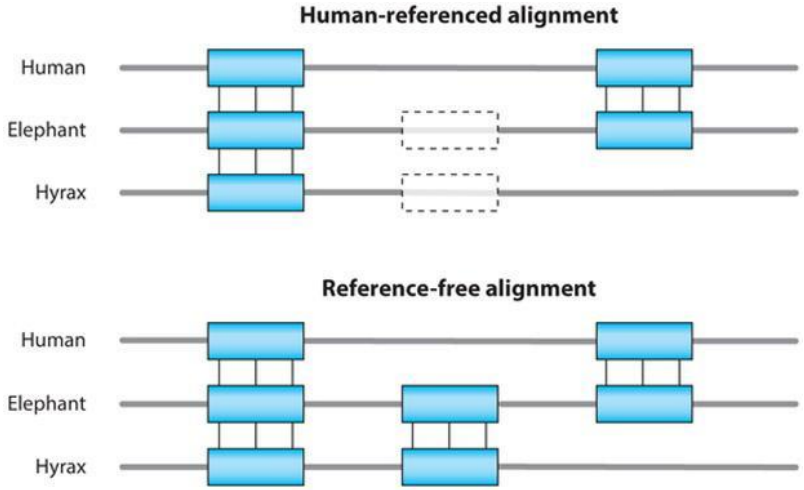
- Studying biological similarities and differences, evolutionary relationships
- Organisms share genetic material that is conserved between them
- Examples:
  - Gene expression variation
  - Bacteria resistance to antibiotics
  - Molecular epidemiology
  - Phylogenetics
  - Estimation of mutation rates
  - Plant populations studies
  - Genome evolution and function
  - Intra and interspecies genes comparison
  - Comparing healthy or normal vs different microbiomes



Workflow: #1-Alignment

Perform alignment

- **Pairwise or multiple alignments**
  - base on single reference genome
  - Reference-free
- Genome aligners have 2 stages:
  - **Filtering** (generating local alignments and filtering to remove false-positive alignments, and identify homologous rearrangement-free regions)
  - **Refinement** (homologous regions aligned with collinear aligner, constructing graph representation of alignment using various heuristics)
- Look for **orthology** (sequence that shares common ancestry)
- Analyze extent of relatedness



Main alignment methods

Pairwise genome alignment

Table 1 Pairwise genome alignment tools

Program	Year (Reference)	Description
MUMmer	1999 (4, 117)	Fast aligner relying on maximal unique matches from a query sequence to a reference sequence; recent versions remove the colinearity restriction of the first version and improve the speed
Chains and nets	2003 (29)	Combines fragmented local alignments into larger, high-scoring chains, which are arranged into hierarchical nets representing rearrangements
Shuffle-LAGAN	2003 (27)	A glocal (global + local) aligner that is less restrictive than global alignment but still enforces monotonicity of the blocks relative to one sequence

Main alignment methods

Multiple genome alignment

Table 2 Popular and/or historically important multiple genome alignment tools

Program	Year (Reference)	Reference-bias	Single-copy	Description
TBA	2004 (8)		✓	Collinear multiple aligner (using MULTIZ internally) that produces a collection of partially ordered threaded blocksets
Mugsy	2011 (32)			Uses a graph-based method to segment the alignment problem into locally collinear blocks: small subregions with no local rearrangements, which are fed into a collinear multiple aligner
MULTIZ (autoMZ)	2004 (8)	✓	✓	Multiple alignment based on pairwise alignment from every genome to a single reference
ABA	2004 (14)			Aligner based on the concept of A-Bruijn graphs
EPO	2008 (13, 37)	*		Graph-based aligner that allows duplications and optionally produces ancestral reconstructions
VISTA-LAGAN (SuperMap)	2009 (36)			Progressive aligner based on Shuffle-LAGAN (27)
Mauve	2004 (25)		✓	Finds maximal unique matches present in every input species, then attempts to remove small matches that cause rearrangements that disrupt collinearity
progressiveMauve	2010 (31)	✓		Progressive aligner that attempts to remove anchors causing small rearrangements by optimizing a breakpoint-weighted score
Cactus	2011 (40)			Graph-based aligner that attempts to remove anchors representing small rearrangements

Workflow: #2: Annotation

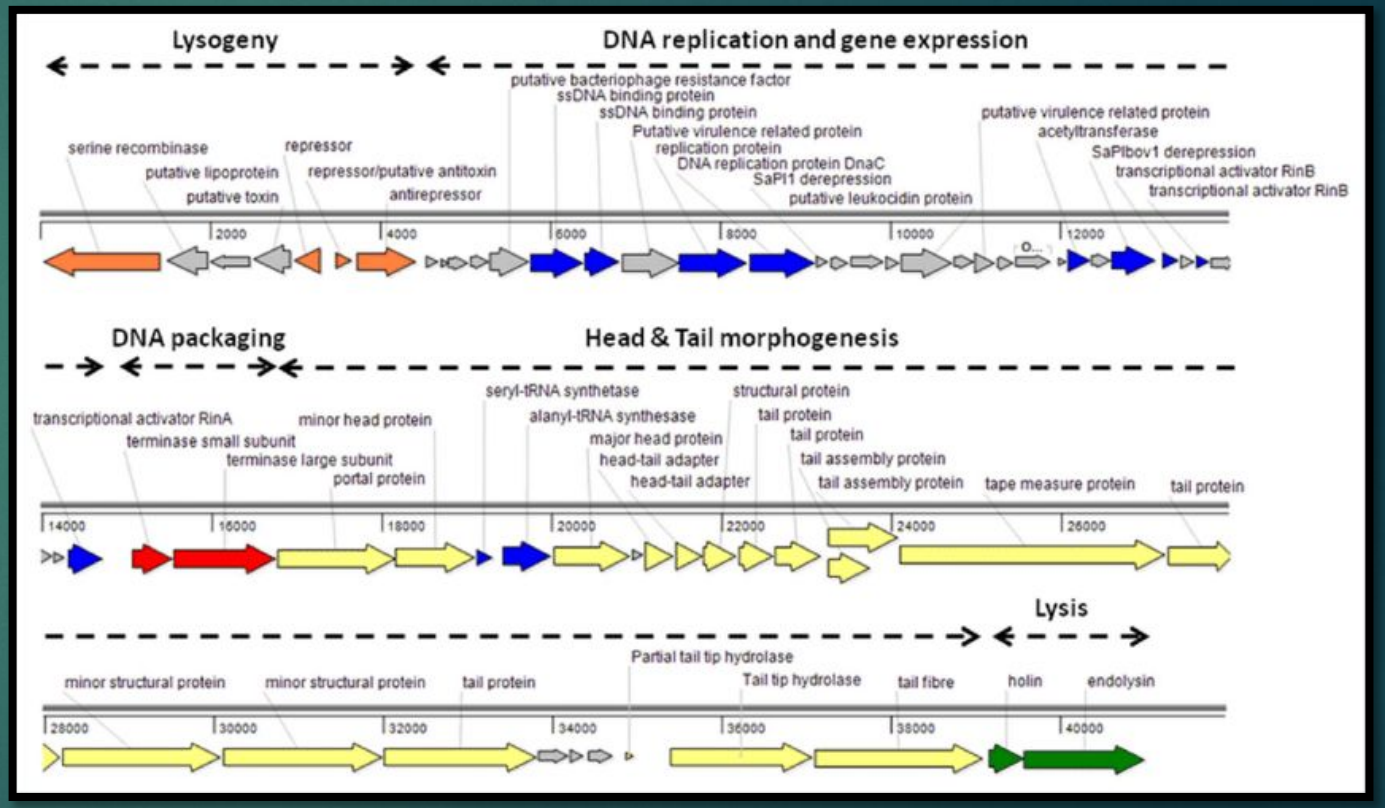
- **Genome annotation: The process of finding functional elements in a genome assembly**
  - Functional elements include:
    - Protein coding genes
    - Non-coding transcripts
    - Chromatic configuration
    - Dnase hypersensitivity
    - CpG islands
    - Population variation

- **Steps of automatic annotation of genomes**

1. Ab initio prediction
   - Computational prediction of exon-intron structure using statistical models
2. Sequence alignment-based approaches
   - Mapping ESTs (expressed sequence tags)
   - cDNA
   - or protein sequences onto an assembled sequence

Comparative annotation

- Annotation is the central aspect of using genomic data for a specific purpose
- Rapid improvements in sequencing technologies provide high quality genome assemblies at affordable prices
- **3 main types of annotations:**
  1. **Sequence-based**
  2. **Transcriptome Evidence-based**
  3. **Transcript Projection**



Comparative annotation: #1: Sequence based

- Combines established single-genome gene prediction approaches with data obtained through genomic alignments to improve gene prediction
- Needs pairwise alignments

Table 3 Overview of comparative annotation tools

Program	Year (Reference)	Description
ROSETTA	2000 (110)	Uses pairwise genomic alignments to find regions of homology; incorporates a splice junction and exon length model
SGP-1/-2	2001 (61)	Uses pairwise genomic alignments to find syntenic loci; evaluates a coding and splice model in these loci
TWINSCAN	2003 (60)	Uses local alignments between a target genome and a reference (informant) genome to identify regions of conservation
SLAM	2003 (62)	Treats two alignments in a symmetric way, predicting pairs of transcripts
EvoGene	2003 (111)	Phylogenetic HMM that performs ab initio prediction of genes across a multiple-sequence alignment (more than two genomes), making use of phylogenetic information
ExoniPhy	2004 (112)	Phylogenetic HMM that performs ab initio predictions across a multiple-sequence alignment
DOGFISH	2006 (113)	Two-step program that combines a classifier that scores potential splice sites using a multiple-sequence alignment and an ab initio gene predictor that makes use of the scores from the classifier to predict gene structures
N-SCAN	2006 (65)	Extends the TWINSCAN model to *N* genomes
CONTRAST	2007 (68)	Uses a combination of SVM and CRF predictors, providing a big boost over traditional HMMs
DOUBLESCAN	2002 (118)	Uses a pair HMM to simultaneously predict gene structures and conservation in two aligned sequences

Abbreviation: CRF, conditional random field; HMM, hidden Markov model; SVM, support vector machine.

Comparative annotation: #2: Transcriptome Evidence-based

- **Mapping transcriptome data to assembly**
  - The transcriptome data can be used in a comparative way because many species of interest have limited transctipome data but are closely related to well-annotated species.
  - Examples: mouse versus rat and human versus apes
- **Mapping protein sequences**
  - They are more robust across long phylogenetic distances

Table 4 Overview of gene prediction tools that incorporate transcriptome data

Program	Year (Reference)	Description
GeneWise	2004 (75)	HMM-based gene prediction tool using extrinsic evidence; MAKER2 can make use of it
N-SCAN-EST	2006 (74)	HMM-based gene prediction tool that makes use of EST and genomic alignments, incorporating phylogenetic information
AUGUSTUS	2004 (76, 77)	CRF-based gene prediction tool with many modes; features are still being added; can perform ab initio gene prediction as well as incorporate extrinsic evidence; has the ability to provide nonlinear weights to various types of evidence
EVM	2008 (114, 115)	A chooser algorithm that combines previously predicted gene sets with extrinsic information to construct consensus gene sets
PASA	2003 (116)	Uses alignments of cDNA, EST, or RNA-seq to predict gene structures, including alternative splice events
MAKER2	2008 (55, 78)	An all-in-one pipeline that runs programs including AUGUSTUS and GeneWise with extrinsic information such as RNA-seq or protein sequences to both predict annotations and construct a gene set

Abbreviation: cDNA, complementary DNA; CRF, conditional random field; EST, expressed sequence tag; HMM, hidden Markov model; RNA-seq, RNA sequencing.

Comparative annotation: #3: transcript projection

- works by projecting the coordinates of an existing annotation in one genome to another genome
- uses high-quality annotations in well-studied organisms to annotate diverse transcripts in related genomes

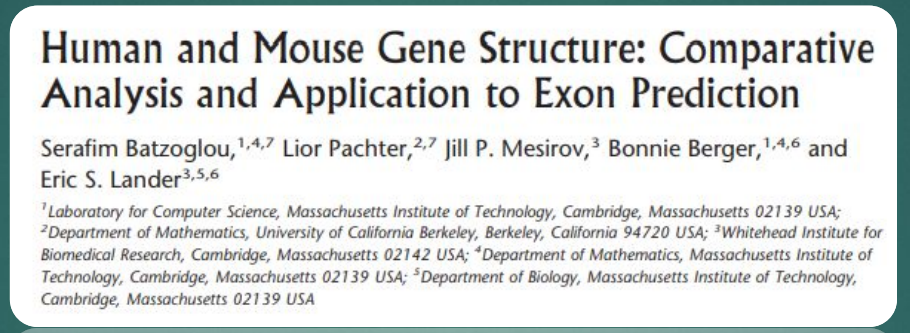
Table 5 Overview of transcript projection tools

Program	Year (Reference)	Description
Projector	2004 (85)	Similar to DOUBLESCAN but extends the model to make use of annotation information on one sequence to inform the other; works better than GENEWISE over long branch lengths
AIR	2005 (86)	Integrates multiple forms of extrinsic evidence to perform alternative splice junction prediction
transMap	2007 (70, 87)	Uses whole-genome alignments to project existing annotations from one genome to one or more other genomes
CESAR	2016 (84)	Uses a HMM to adjust splice sites in whole-genome alignments, improving transcript projections

Abbreviations: AIR, Annotation Integrated Resource; CESAR, Coding Exon-Structure Aware Realigner; HMM, hidden Markov model.

Mini review of 1st comparative genomics paper

Mini review of early paper: "Human and Mouse Gene Structure: Comparative Analysis and Application to Exon Prediction" (Batzoglou et al 2000)



Human and Mouse Gene Structure: Comparative Analysis and Application to Exon Prediction

Serafim Batzoglou,[1,4,7] Lior Pachter,[2,7] Jill P. Mesirov,[3] Bonnie Berger,[1,4,6] and Eric S. Lander[3,5,6]

[1]Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA; [2]Department of Mathematics, University of California Berkeley, Berkeley, California 94720 USA; [3]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142 USA; [4]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA; [5]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA

Objectives:

- 1st cross-species sequence comparison: simultaneously analyzing homologous loci from 2 related species
- Accurately defining coding exons by comparison of syntenic (occurring on the same chromosome) human and mouse genomic sequences
- Developed automatic approach to exon recognition by using cross-species sequence comparison to identify and align relevant regions and search for presence of exonic features at corresponding positions in both species

Mini review of early paper: "Human and Mouse Gene Structure: Comparative Analysis and Application to Exon Prediction" (Batzoglou et al 2000)

Approach:

- Systematic comparison of genomic structure of 117 orthologous gene pairs from human and mouse to understand extent of conservation of the **number**, **length**, and **sequence** of exons and introns.
- Development of algorithms for cross-species gene recognition, called "**GLASS**", and "**ROSETTA**", an exon identifying program
  - **GLASS** makes good global alignments of large genomic regions by using hierarchical alignment approach
  - **ROSETTA** identified coding exons, with 95% sensitivity and 97% specificity

Mini review of early paper: "Human and Mouse Gene Structure: Comparative Analysis and Application to Exon Prediction" (Batzoglou et al 2000)

Results

- **Exons:**
  - Revealed great extent of evolutionary conservation
  - # of exons was identical for 95% of genes studied
  - Lengths of exons also strongly conserved, identical in 73%
- **Sequence similarity:**
  - Coding regions 85% similar
  - Introns differed (35% similarity), human introns larger
  - Among genes: 88% at DNA level and 100% identity at amino acid level

Table 1. Comparative Analysis of Human and Mouse Loci

		Total	5' NC	C.ex.	3'NC	Intron	Alignments of Regions
1	HSCKIIBE	5917	340	648	140	3665	329 611 11 72 668 103 569 116 446 76 145 190 327 91 140
	MMGMCK2B	7874	332	648	140	4129	321 593 11 72 588 103 1160 116 387 76 110 190 1271 91 140
	Casein kinase II subunit beta gene	57.51	66.07	93.06 98.61	77.14	40.89	66.15 56.31 63.64 94.44 100 -+ 31.79 95.15 99.03 ++ 36.13 89.66 98.28 ++ 50.9 92.11 98.68 ++ 54.69 93.16 97.89 ++ 27.41 94.51 98.9 ++ 77.14
2	HUMSAACT	3778	103	1134	253	1345	91 869 12 129 106 325 127 162 79 192 86 182 78 144 253
	MUSACASA	4007	70	1134	242	1498	58 981 12 129 98 325 140 162 93 192 132 182 74 144 242
	Skeletal alpha-actin gene	59.53	41.34	90.3 99.21	53.7	31.75	33.6 26.12 100 89.15 100 -+ 44.12 88.92 99.69 ++ 38.7 89.51 98.15 ++ 48.35 90.1 98.44 ++ 51.36 92.31 98.9 ++ 34.21 93.06 100 ++ 53.7
3	HSH4EHIS	859	NA	312	NA	NA	312
	MMHIS412	637	NA	312	NA	NA	312
	Histone H4 gene	57.61	NA	89.42 100	NA	NA	89.42 100 ++
4	HSU12202	4942	36	393	81	3521	36 3 1443 66 93 210 1360 111 485 3 19 600 62 - -
	MMMRPS24	5499	46	396	200	4597	46 3 930 66 91 210 1265 111 558 6 12 1358 20 325 84
	Ribosomal protein S24 gene	34.42	41.46	88.55 100	4.539	16.9829	41.46 100 100 X +23.76 84.85 100 ++ 43.48 89.05 100 ++ 16.64 90.99 100 ++ 25.75 66.7 100 X 19.35 29.72 0 0 0
5	HUMHIS4	1098	NA	312	NA	NA	312
	MUSHIST4	968	NA	312	NA	NA	312
	Histone H4 gene	48.86	NA	87.18 100	NA	NA	87.18 100 -+
6	HSHISH3	698	NA	411	NA	NA	411
	MMHIST31	592	NA	411	NA	NA	411
	Histone H3 gene	72.13	NA	85.89 100	NA	NA	85.89 100 ++
7	HSHSC70	5408	83	1941	NA	2380	78 730 5 205 322 206 324 153 87 556 211 203 228 199 147 233 331 186 NA
	MMU73744	4270	65	1941	85	1840	60 566 5 205 308 206 115 153 84 556 212 203 222 199 95 233 238 186 65
	Hsc70 gene for heat shock cognate protein	61.62	52.47	88.97 99.23	NA	30.23	50.7 26.1 80 86.34 99.51 -+ 22.66 89.32 99.03 ++ 13.67 90.2 100 ++ 47.95 89.57 99.82 ++ 43.5 92.12 97.54 ++ 52.44 84.92 99.5 ++ 14.05 89.7 97.85 ++ 41.48 88.71 100 ++ NA
8	HUMNOCT	4878	NA	1332	NA	NA	1332
	MUSPOUDOMB	3864	NA	1338	NA	NA	1338
	POU domain transcription factor	71.58	NA	93.84 99.1	NA	NA	93.63 99.1 ++
9	HUMTROC	4567	27	486	173	2244	27 24 1466 31 230 147 245 115 216 137 84 32 173
	MUSCTNC	4194	44	486	173	2687	44 24 1418 31 226 147 354 115 609 137 87 32 173
	Slow twitch skeletal nuscle cardiac troponin	54.41	62	89.92 96.91	59	48.90	62 91.67 100 ++ 54.09 83.87 96.77 ++ 47.37 91.16 97.96 ++ 32.23 92.17 96.52 ++ 33.7 86.86 96.35 ++ 45.61 93.75 93.75 ++ 59
10	HSINT1G	4522	NA	1113	NA	1877	104 713 254 702 266 462 489
	MUSINT1A	5607	NA	1113	NA	1599	104 569 254 573 266 457 489
	Int-1 mammary oncogene	65.33	NA	91.11 98.38	NA	46.21	90.38 41.97 88.98 49.57 91.35 47.66 92.23 89.42 98.03 ++ 99.25 100 -+ ++ ++

: Coding Exons
: Noncoding Exons
: Introns

Mini review of early paper: "Human and Mouse Gene Structure: Comparative Analysis and Application to Exon Prediction" (Batzoglou et al 2000)

Results

- **GLASS** (Global sequence Alignment) analysis
- Output from **GLASS** is then used in **ROSETTA**, for gene recognition

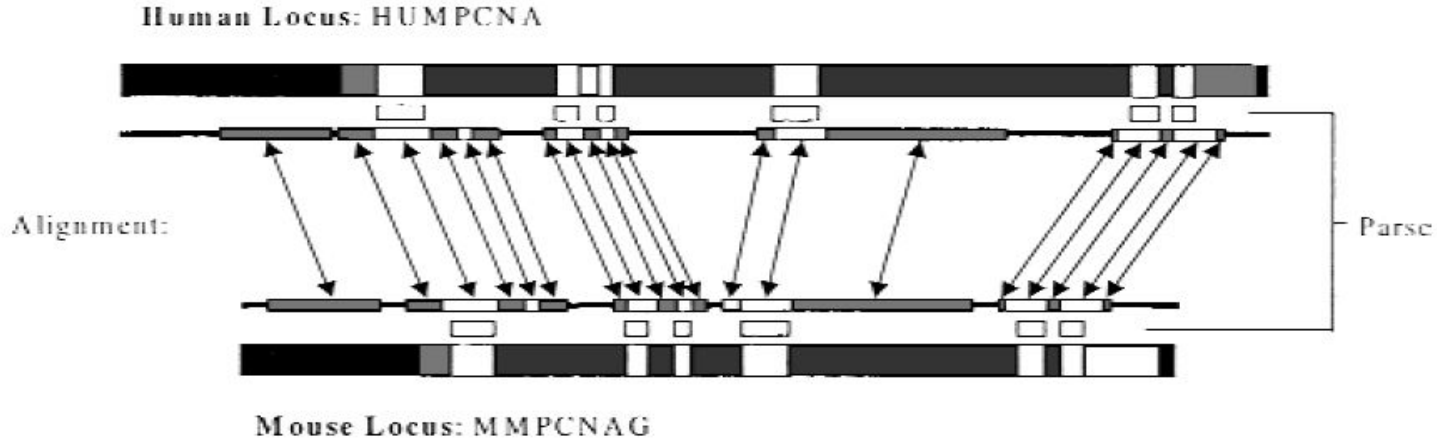


Figure 1 Regions of the human and mouse homologous genes: Coding exons (white), noncoding exons (gray}, introns (dark gray), and intergenic regions (black). Corresponding strong (white) and weak (gray) alignment regions of GLASS are shown connected with arrows. Dark lines connecting the alignment regions denote very weak or no alignment. The predicted coding regions of ROSETTA in human, and the corresponding regins in mouse, are shown (white) between the genes and the alignment regions.

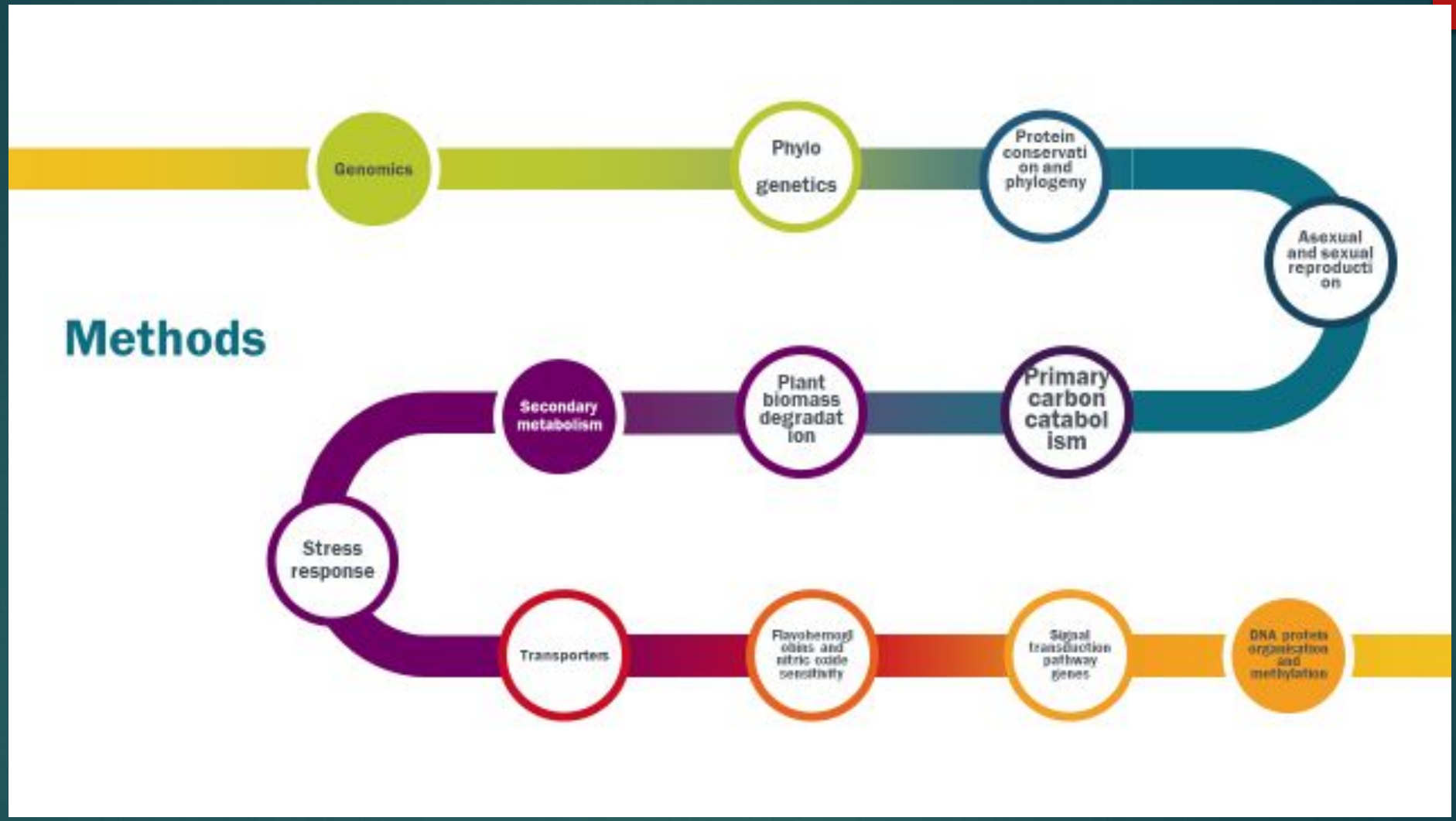
Review #1

Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus Aspergillus.

de Vries, R. P., Riley, R., Wiebenga, A., Aguilar-Osorio, G., Amillis, S., Uchima, C. A., … Grigoriev, I. V. (2017). Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus Aspergillus. *Genome biology*, *18*(1), 28. doi:10.1186/s13059-017-1151-0

Importance of Aspergillus

- The fungal genus Aspergillus is of critical importance to humankind.
- Species include those with industrial applications,
- Important pathogens of humans, animals and crops,
- A source of potent carcinogenic contaminants of food,
- Important genetic model.
- The genome sequences of eight aspergilli have already been explored to investigate aspects of fungal biology, raising questions about evolution and specialization within this genus.


Methods
Genomics
Phylo genetics
Protein conservation and phylogeny
Asexual and sexual reproduction
Primary carbon catabolism
Plant biomass degradation
Secondary metabolism
Stress response
Transporters
Flavohemoglobins and nitric oxide sensitivity
Signal transduction pathway genes
DNA protein organization and methylation

General comparison of Genomics and Phylogenetics

- Analysis and comparison of genomes from different species
- Purposes
  - to gain a better understanding of how species have evolved
  - to determine the function of genes and non-coding regions of the genome
- The functions of human genes and other DNA regions often are revealed by studying their parallels in nonhumans.
  - Researchers have learned a great deal about the function of human genes by examining their counterparts in simpler model organisms such as the mouse.

Features looked at when comparing genomes:

- sequence similarity
- gene location
- length and number of coding regions within genes
- amount of non-coding DNA in each genome
- highly conserved regions maintained in organisms

- Computer programs that can line up multiple genomes and look for regions of similarity among them are used.
- Many of these sequence-similarity tools, such as BLAST, are accessible to the public over the Internet.

Phylogenetics

- Study of evolutionary relationships (sequences / species)
- Infer evolutionary relationship from shared features
- Phylogeny
  - Relationship between organisms with common ancestor
- Phylogenetic tree
  - Graph representing evolutionary history of sequences / species

Phylogenetics cont'd

- Premise
  - Members sharing common evolutionary history
- (i.e., common ancestor) are more related to each other
  - Can infer evolutionary relationship from shared features
- Long history of phylogenetics
  - Historically - based on analysis of observable features
- (e.g., morphology, behavior, geographical distribution)
  - Now - mostly analysis of DNA / RNA / amino acid sequences

Phylogenetics

- Goals
  - Understand relationship of sequence to similar sequences
  - Construct phylogenetic tree representing evolutionary history
- Motivation / application
  - Identify closely related families
    - Use phylogenetic relationships to predict gene function
  - Follow changes in rapidly evolving species (e.g., viruses)
    - Analysis can reveal which genes are under selection
    - Provide epidemiology for tracking infections & vectors
- Relationship to multiple sequence alignment (MSA)
  - Alignment of sequences should take evolution into account
  - More precise phylogenetic relationships Improved MSA
  - CLUTALW (http://www.ebi.ac.uk/clustalw/), a popular MSA program, can produce alignment that is then used to build phylogenetic tree.

Results & Discussion

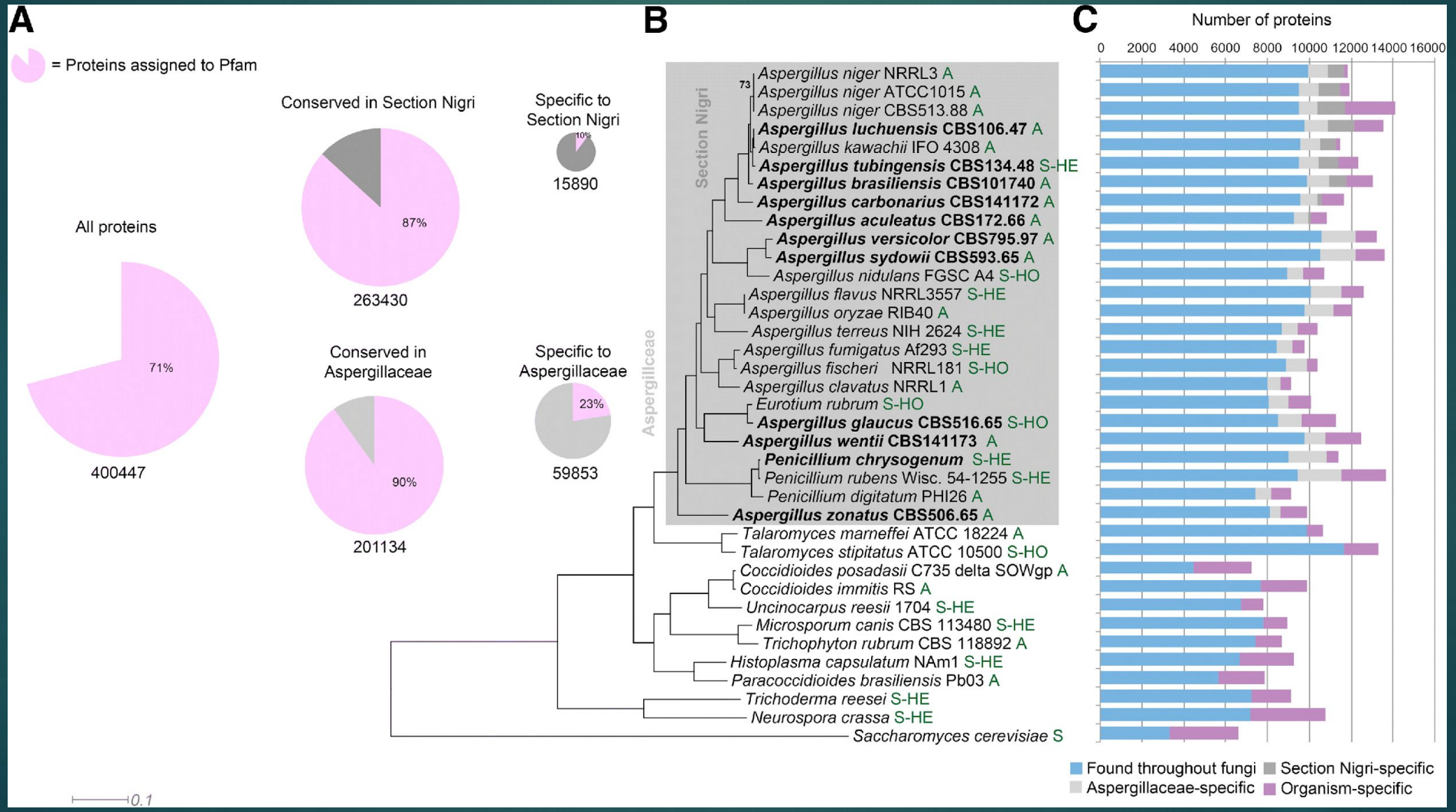
General comparison of the genomes and phylogenomics

Comparison of genome features of 34 ascomycete genomes

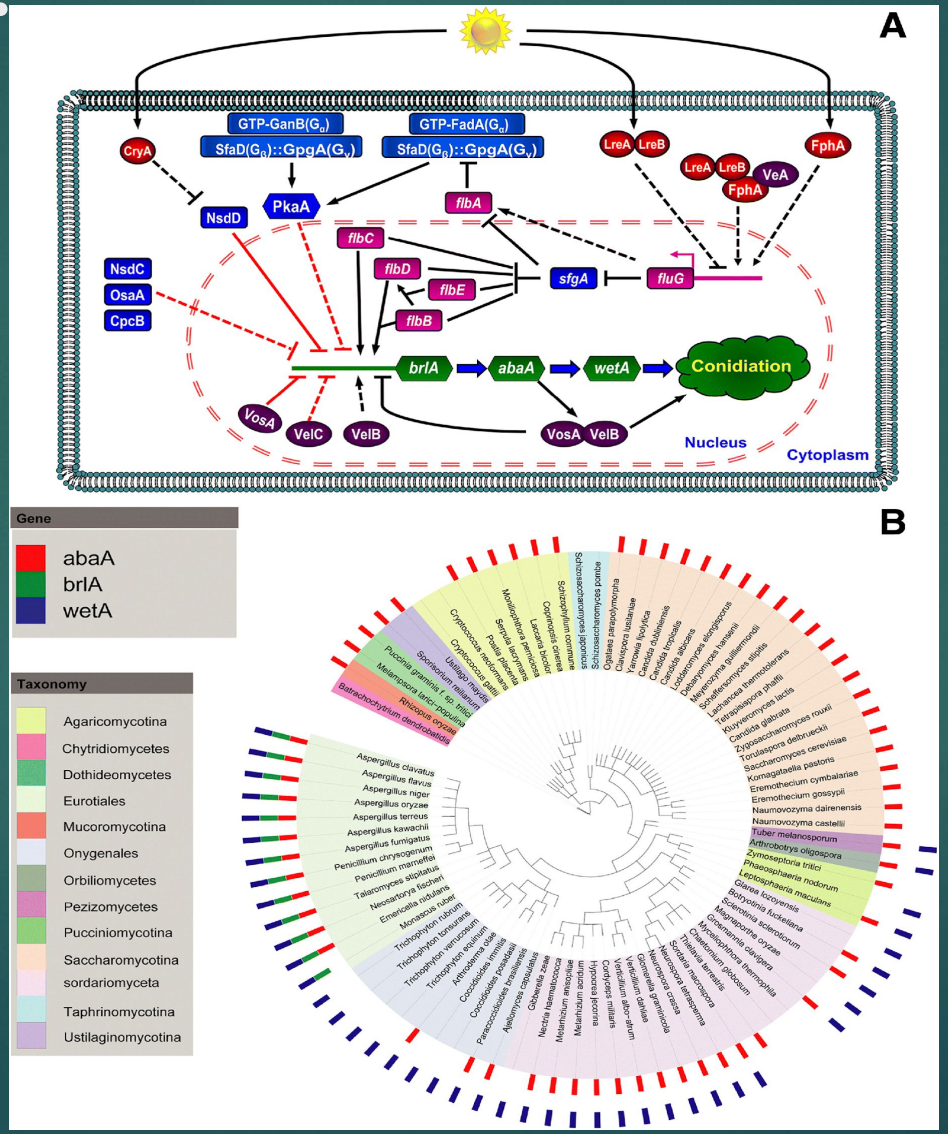
Species	Strain	Class	Order	Section	Genome size (Mb)	GC %	Contigs (*n*)	Reference
Aspergillus niger	ATCC1015	Eurotiomycetes	Eurotiales	*Nigri*	35	50.3	24	[7]
Aspergillus niger	CBS513.88	Eurotiomycetes	Eurotiales	*Nigri*	34	50.5	470	[10]
Aspergillus luchuensis	**CBS106.47**	**Eurotiomycetes**	**Eurotiales**	***Nigri***	**37**	**49.1**	**318**	**This study**
Aspergillus tubingensis	**CBS134.48**	**Eurotiomycetes**	**Eurotiales**	***Nigri***	**35**	**49.2**	**87**	**This study**
Aspergillus brasiliensis	**CBS101740**	**Eurotiomycetes**	**Eurotiales**	***Nigri***	**36**	**50.5**	**290**	**This study**
Aspergillus carbonarius	**CBS141172**	**Eurotiomycetes**	**Eurotiales**	***Nigri***	**36**	**51.7**	**1346**	**This study**
Aspergillus aculeatus	**CBS172.66**	**Eurotiomycetes**	**Eurotiales**	***Nigri***	**35**	**50.9**	**851**	**This study**
Aspergillus versicolor	**CBS795.97**	**Eurotiomycetes**	**Eurotiales**	***Nidulantes***	**33**	**50.1**	**58**	**This study**
Aspergillus sydowii	**CBS593.65**	**Eurotiomycetes**	**Eurotiales**	***Nidulantes***	**34**	**50.0**	**133**	**This study**

The genome sequences were generated at the Joint Genome Institute (JGI) and compared to other fungal genomes in JGI's MycoCosm database with respect to genome size, structure, and gene content

Genome overview of *Aspergillus* and comparative species



Regulatory pathway of asexual sporulation in *A. nidulans*.



Conclusions

- Genome sequences for 10 novel highly diverse Aspergillus species generated
- Comparison in detail to sister and more distant genera.
- Comparative studies of key aspects of fungal biology, including:
  - primary and secondary metabolism,
  - stress response,
  - biomass degradation,
  - and signal transduction,

  revealed both conservation and diversity among the species.

Conclusions cont'd

Observed genomic differences were validated with experimental studies. This revealed several highlights, such as:

- potential for sex in asexual species,
- organic acid production genes
- being a key feature of black aspergilli,
- alternative approaches for degrading plant biomass,
- and indications for the genetic basis of stress response.

A genome-wide phylogenetic analysis demonstrated in detail the relationship of the newly genome sequenced species with other aspergilli.

Review #2

Comparative analysis of an *mcr-4* *Salmonella enterica* subsp. *enterica* monophasic variant of human and animal origin

ALESSANDRA CARATTOLI EDOARDO CARRETTO FLAVIA BROVARONE MARIO SARTI LAURA VILLA

JOURNAL OF ANTIMICROBIAL CHEMOTHERAPY, VOLUME 73, ISSUE 12, DECEMBER 2018, PAGES 3332–3335

Objectives

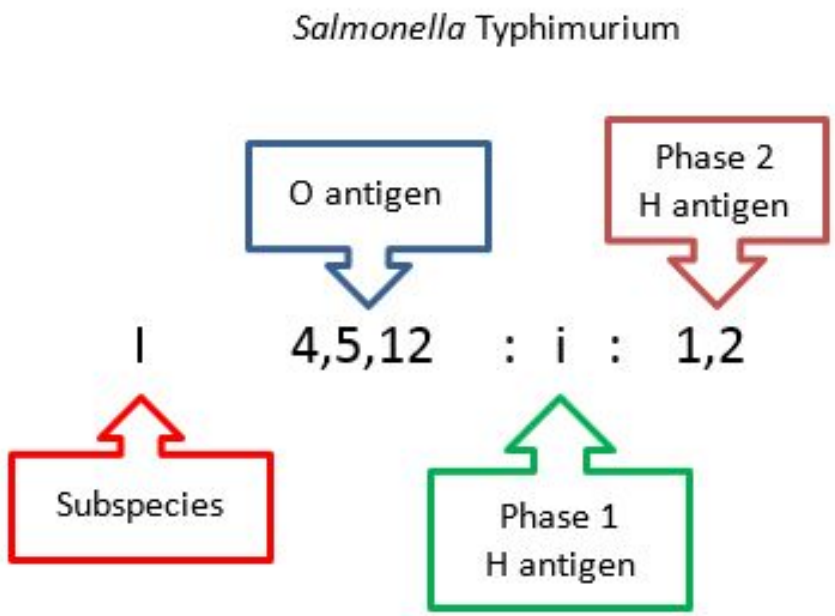
- To compare the *mcr-4*-positive *Salmonella enterica* monophasic variant from two Italian patients affected by gastroenteritis with the first *mcr-4*-positive *Salmonella* isolate identified in a pig at slaughter in Italy.

Introduction

- *Salmonella enterica* monophasic variant 4,[5],12:i:-
- Colistin is an antibiotic usually used in animals but recently made a return in human treatments due to emerging antibiotic resistant bacteria
- There are two described colistin resistance mechanisms:

- Chromosomal mutations involving the LPS; and

- Acquisition of *mcr* (mobile colistin resistance) genes

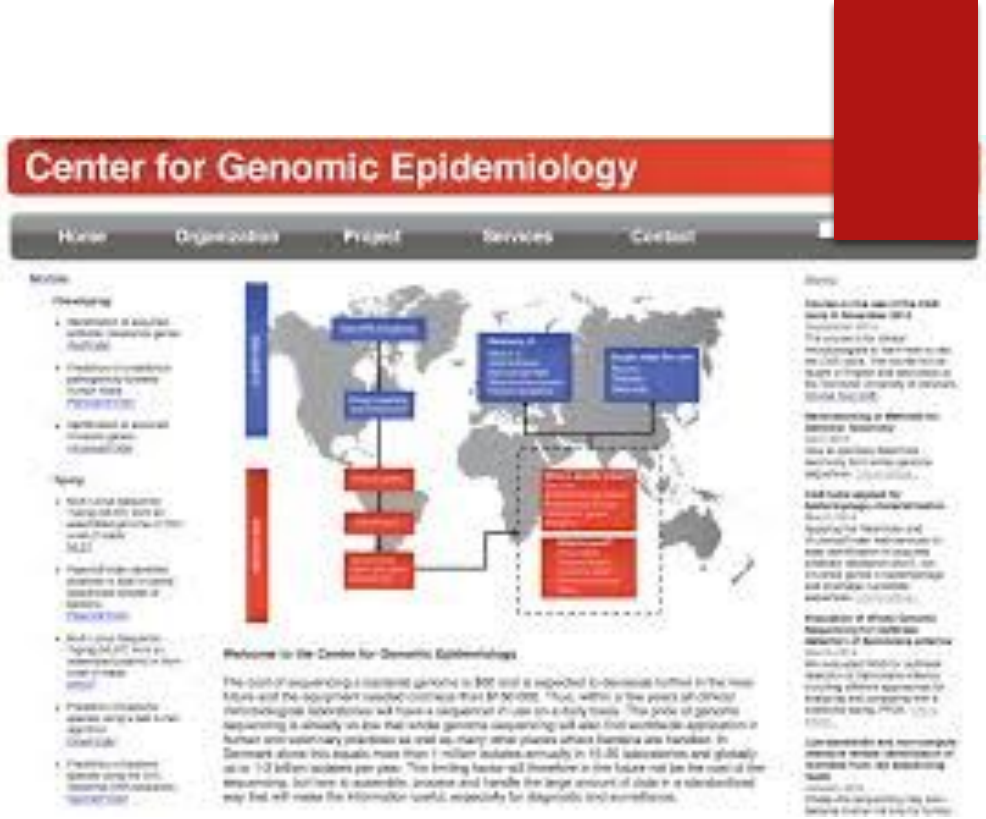


Introduction

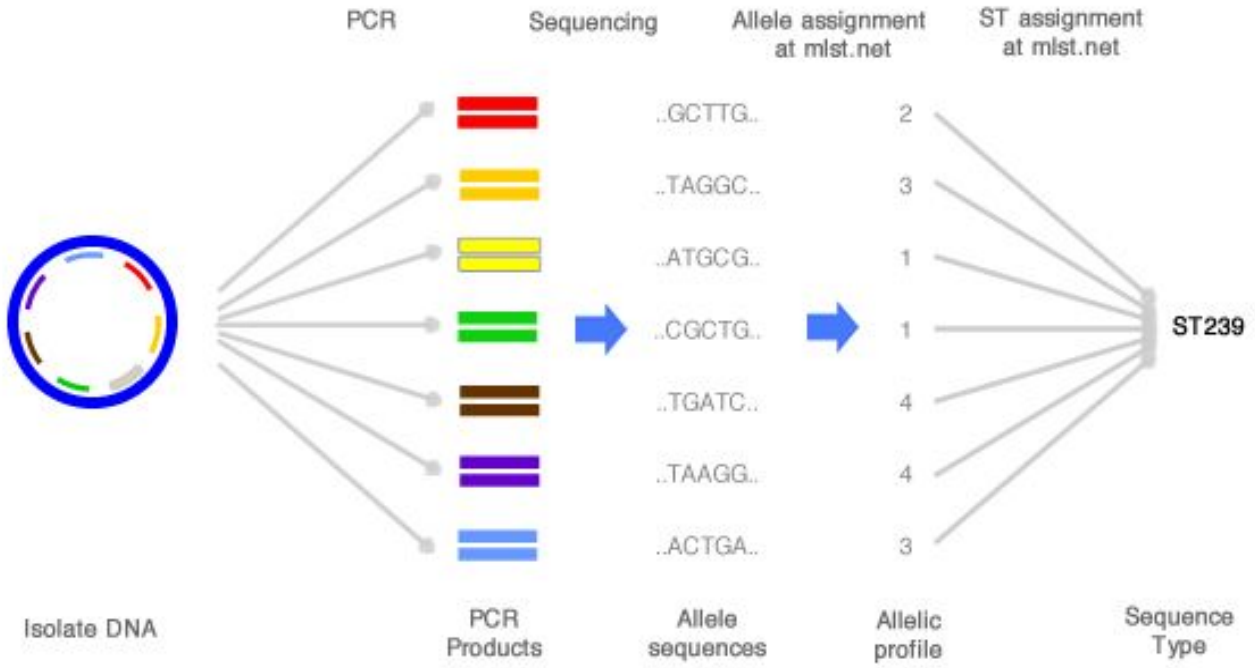
- There are 5 types of *mcr*: *mcr-1, mcr-2, mcr-3,* ***mcr-4****, mcr-5*
- *mcr-4.1* found in *Salmonella* isolated from a pig slaughtered in 2013
- *mcr-4.2* found in *Salmonella* isolated from human gastroenteritis patients in 2016

Methods

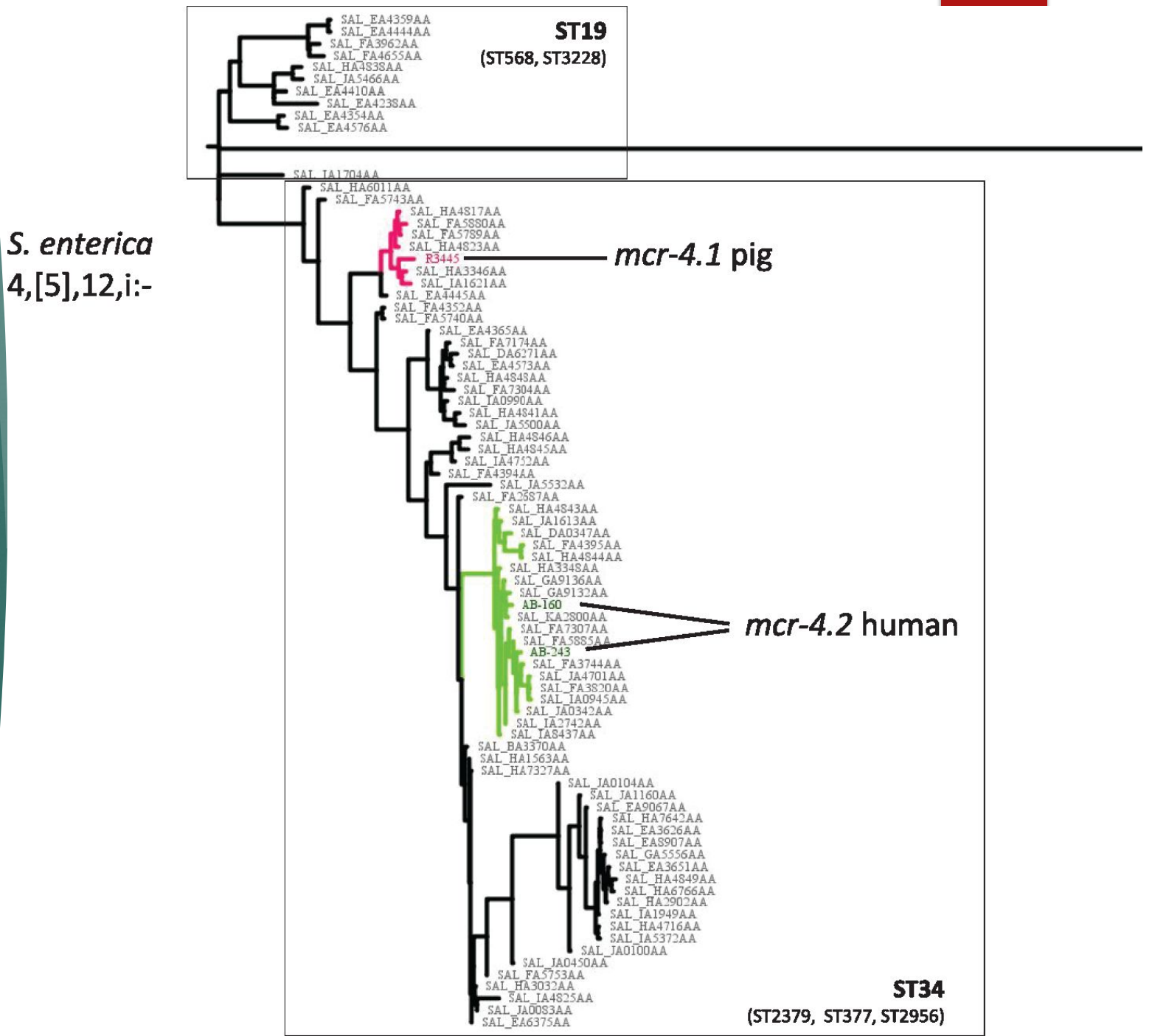
- WGS of the *Salmonella* isolates was performed using a MiSeq instrument (Illumina) and paired-end libraries generated with Nextera XT (Illumina)
- Draft genome sequences annotated with RAST server and antimicrobial resistance genes detected using ResFinder and PlasmidFinder



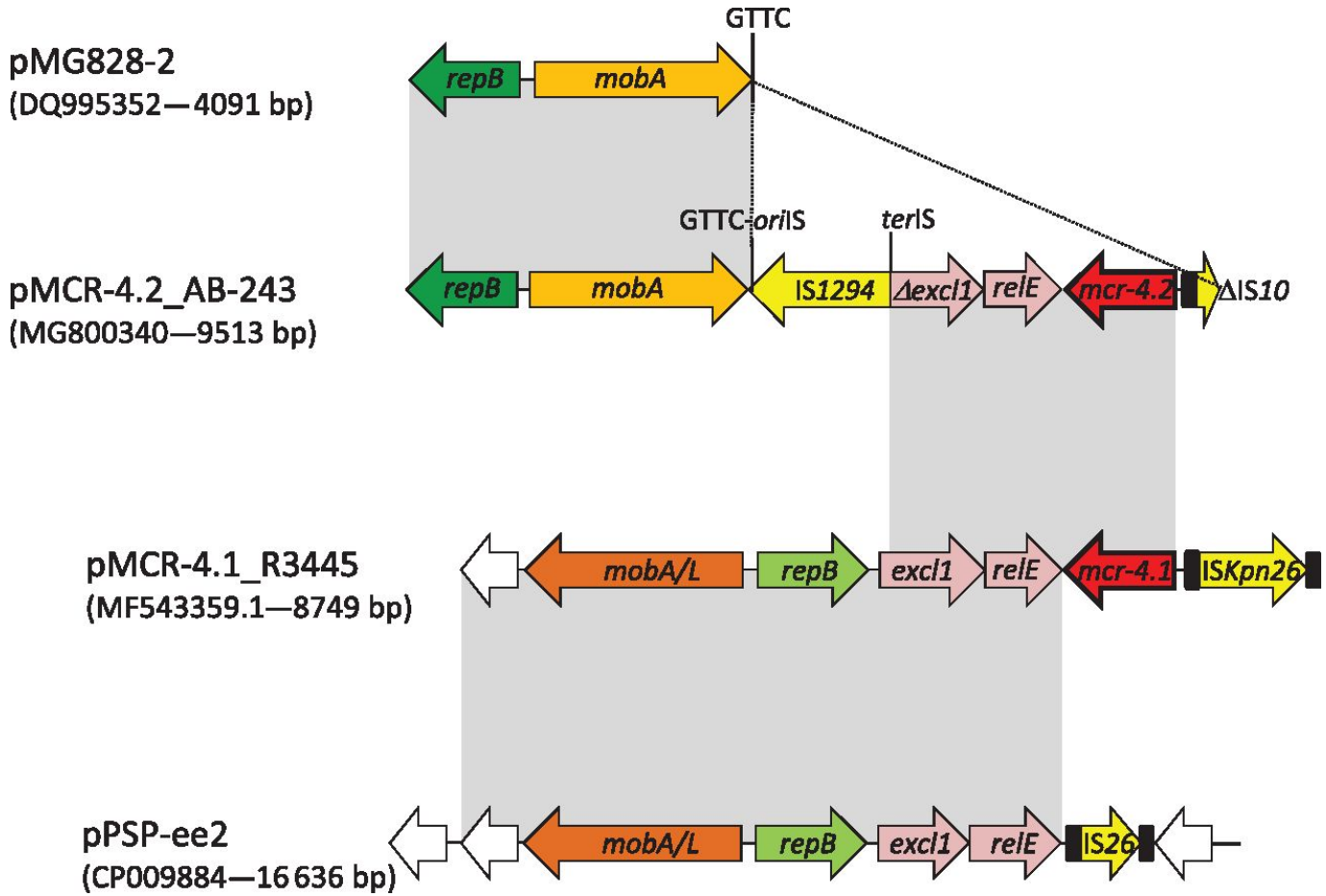
- The 82 *Salmonella* genomes downloaded from the EnteroBase *Salmonella* database + isolates of swine and human origin used for the phylogenetic tree based on MLST STs



- Phylogenetic maximum-likelihood tree analysis was done on FigTree v1.4.3 software using the SNP analysis, comparing genomes of the mcr-4-positive isolates with the database
- Complete sequences of plasmids carrying mcr-4.2 were obtained and compared
- Comparative genomics demonstrated that the Salmonella of swine origin did not cluster with the isolates of human origin



- Comparative linear maps of plasmids and their close relatives
- The *mcr-4.2* gene variant identified in the *Salmonella* of human origin was located on a ColE-like plasmid
- This plasmid showed different replication and mobilization genes with respect to those previously described in the ColE plasmid carrying the *mcr-4.1* variant, identified in *Salmonella* of swine origin.

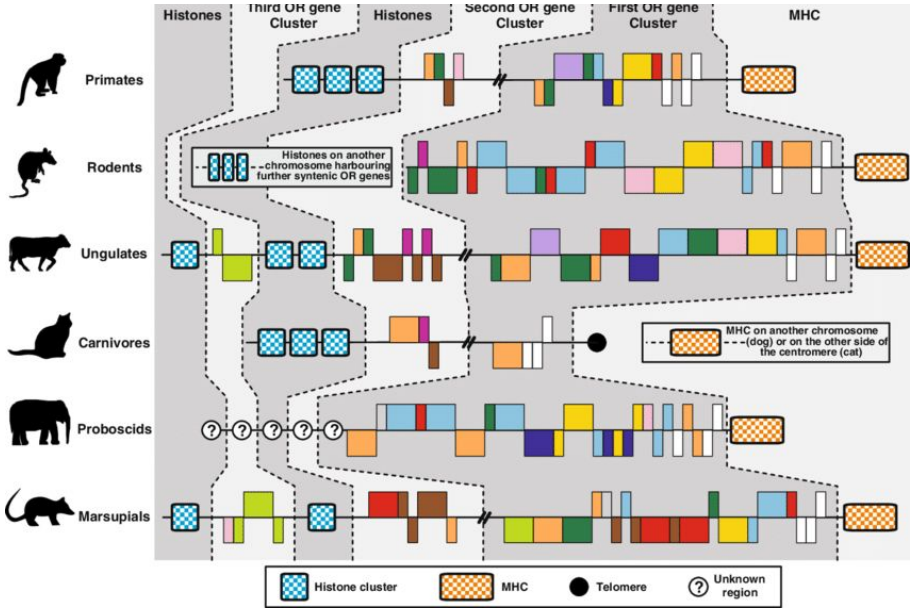


Conclusions

- There were differences found between genomes, plasmids and gene variants demonstrated
- The *mcr-4*-positive *Salmonella* lineage circulating in animals and lineage causing gastroenteritis in humans in Italy were not the same
- There was no horizontal transfer of the same plasmid among *Salmonella* strains of animal and human origin
- The *mcr-4* gene and a fragment of the plasmid identified in the animal strain were mobilized by an IS*1294* into a different plasmid.

Final Summary

- Genome technology relatively new.
  - First human genome in 2003
  - Comparative genomics research took off with first few assembly and alignment technologies
- Many taxonomic levels of organisms compared and phylogenetically analyzed
- Countless applications in the medical, agricultural, animal fields
- 2 broad steps (each have choice of many technologies):
  - Alignment
    - Pairwise, multiple
  - Annotation
    - Sequence based, transcriptome evidence-based or projection



References

- Armstrong, J., Fiddes, I. T., Diekhans, M., & Paten, B. (2018). Whole-Genome Alignment and Comparative Annotation. *Annual review of animal biosciences*, *7*, 41–64. doi:10.1146/annurev-animal-020518-115005
- Batzoglou S, Pachter L, Mesirov JP, Berger B, Lander ES. 2000. Human and mouse gene structure: comparative analysis and application to exon prediction. Genome Res. 10:950–58
- Carttoli A., Carreto F., Brovarone M., Villa L. (2018). Comparative analysis of an mcr-4 Salmonella enterica subsp. enterica monophasic variant of human and animal origin. Journal of antimicrobial chemotherapy, 73(12), 3332-3335
- de Vries, R. P., Riley, R., Wiebenga, A., Aguilar-Osorio, G., Amillis, S., Uchima, C. A., … Grigoriev, I. V. (2017). Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus Aspergillus. *Genome biology*, *18*(1), 28. doi:10.1186/s13059-017-1151-0