

# Microbial Comparative Genomics and Bioconductor Preview

---

By: Elizabeth Helton



# Overview

- Glimpse at Bioconductor

- What is a Genome?

- What is Comparative Genomics?

- Package 'Find my Friends' compared to others

- Why/ How is it used?

- example

- summary

# Bioconductor

---

- Used for analysis, comprehension, and visual aid of genomic data



# Genome

---

-Organism's complete set of DNA, which includes all of its genes and noncoding sequences



# Comparative genomics

---

- Used to compare complete genome sequences of various species
- Able to identify regions of similarity and differences between species
- Used to better understand the structure and function of human genes and come up with new ways to fight diseases

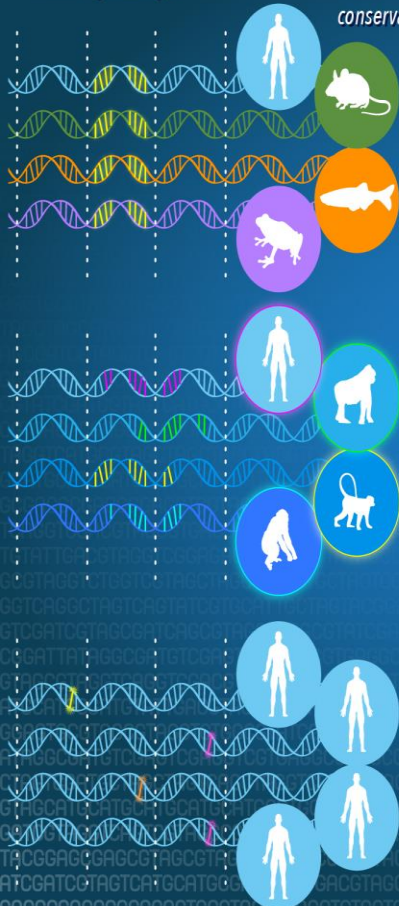


# COMPARATIVE GENOMICS

NHGRI FACT SHEETS

genome.gov

Researchers choose the appropriate time-scale of evolutionary conservation for the question being addressed.



Common features of different organisms such as humans and fish are often encoded within the DNA evolutionarily conserved between them.

Looking at closely related species such as humans and chimpanzees shows which genomic elements are unique to each.

Genetic differences within one species such as our own can reveal variants with a role in disease.

**Table 5. Enhancer trap alleles**

CG#	GENE	LINE	SITE	CHR	Strand	phenotype	T1	Insert	Met	Stop	T2	Insert	Met	Stop	T3	Insert	Met	Stop	T4	Insert	Met	Stop	type			
CG7147	kuz	CB02000	13532398	ZL	+	lethal	RA	1	2	12	RB	1	2	12									4			
CG7892	nmo	CB02015	7938638	3L	+	lethal	RA	1	3	9	RB	1	2	8	RE	1	3	9	RC	1	2	8	4			
CG3758	egp	CB02017	15311950	ZL	+	lethal	RA	1	1	1													4			
CG30403	CG30403	CB02022	16756893	2R	+	lethal	RA	0.5	1	2													3B			
CG9936	skd	CB02029	20843060	4R	+	lethal	RA	1.5	2	13	RE	1.5	2	13									2B			
CG2411	phc	CB02030	3710594	2L	+	lethal	RA	1	1	1													4			
CG17654	Eno	CB02039	1729440	2R	+	lethal	RA	3	2	4	RE	3	3	4	RC	3	3	4	RD	3	3	4	4			
CG3619	DI	CB02040	1515190	2L	+	lethal	RA	0.5	6	6	RB	0.5	1	6									3B			
CG1621	CG1621	CB02042	2553988	2L	+	lethal	RA	2	2	2													3B			
CG12891	CPT1	CB02043	5539025	2R	+	lethal	RA	2	2	9													4			
CG4043	Rrp46	CB02050	6238635	3R	+	lethal	RA	1	1	1													3B			
CG5677	CG5677	CB02054	20046380	3R	-	lethal	RA	1	1	1													4			
CG9062	CG9062	CB02056	6348662	2R	-	lethal	RA	1	1	1					1	1	9						4, 1			
CG10823	CG10823	CB02057	17122275	3R	-	lethal	RA	1	1	1					2	4							3B			
CG9415	Xbp1	CB02061	16187154	2R	+	lethal	RA	1	1	1					1	2							4			
CG30498	boca	CB02070	2622679	2R	-	lethal	RA	1	1	1					1	2							3B			
CG15864	CG15864	CB02071	6076201	3R	+	lethal	RA	1	1	1					1	2							5			
CG13114	Neu3	CB02076	10523056	3R	-	lethal	RA	1	1	1					1	10							3B			
CG8128	CG8128	CB02087	15428524	X	+	lethal	RA	1	1	1					1	10							4			
CG17117	hth	CB02095	6440060	3R	+	semi-lethal	RA	1	1	1	RE	1.5	1	14	RD	1.5	2	8					4			
CG5887	dexam1	CB02105	8269757	3R	+	lethal	RA	1	2	1					1	2	5	RE	1.5	2	5	RB	1.5	2	5	4
CG30497	CG30497	CB02106	2840573	2R	-	lethal	RA	1	1	1					1	2								3B		
CG17342	Lk6	CB02120	7590202	3R	-	lethal	RA	0.5	2	2					1	2								3B		
CG4570	CG4570	CB02124	6682635	3R	+	lethal	RA	1	2	2					1	2								4		
CG1427	CG1427	CB02125	1426793	3R	+	lethal	RA	0.5	2	4					1	2								4		
CG3428	CG3428	CB02131	9445286	3L	-	lethal	RA	1	1	1					1	2								4		
CG14478	CG14478	CB02133	12521410	2R	+	lethal	RA	1	2	2					1	2								4		
CG7820	I(3)B70f	CB02135	8856406	3R	+	lethal	RA	0.5	1	3					1	2								4		
CG2161	Rgs	CB02139	1438299	3R	-	lethal	RA	1	2	2					1	2	8							4		
CG31241	CG31241	CB02140	14081648	3R	+	lethal	RA	0.5	1	1					1	2								4		
CG17328	CG17328	CB02149	16277703	2L	+	lethal	RA	0.5	1	1					1	2								3B		
CG12284	th	CB02150	1	3L	+	lethal	RA	1	1	1					1	2								3B		
CG1837	CG1837	CB02168	11449163	X	+	lethal	RA	0.5	1	3					1	2								3A		

# Find my Friends/ comparison

---

-Author: Thomas Lin Pedersen

-Framework for microbial comparative genomics. Defines a class system for when working with a pangenome datasets. It allows for a transparency to the underlying sequence data while being able to handle massive collections of genomes.

-Defines a set of novel algorithms that make it possible to create a high quality and speedy pangenome sequence.

# Why is it used?

---

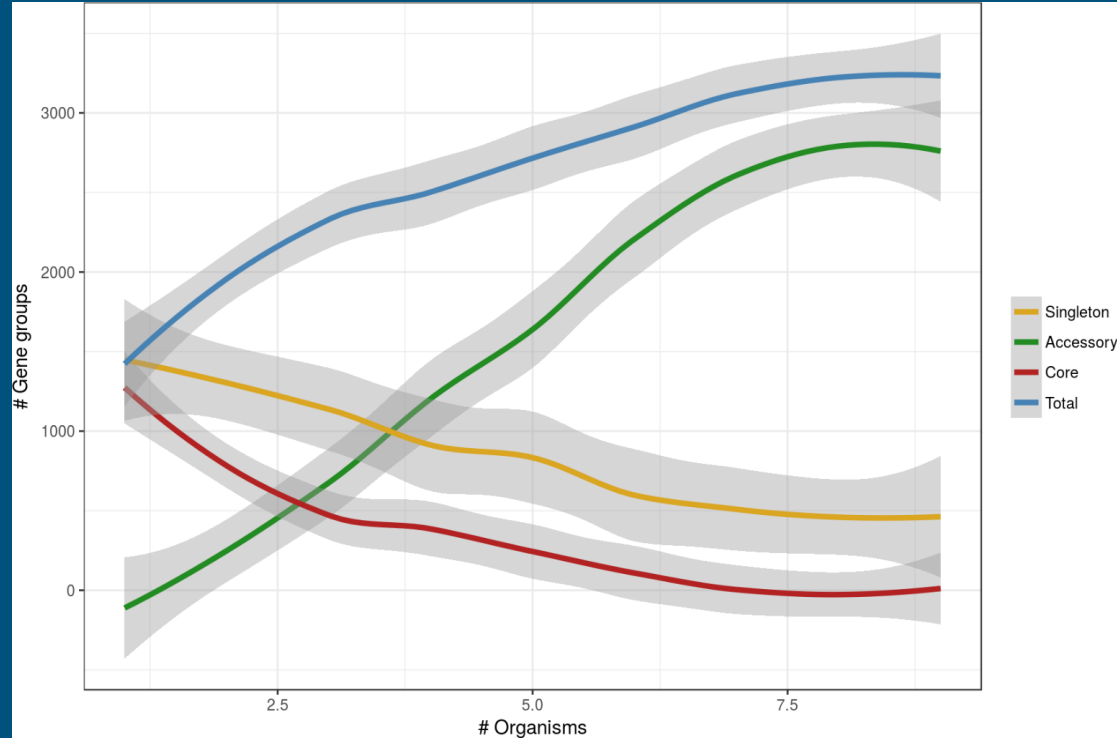
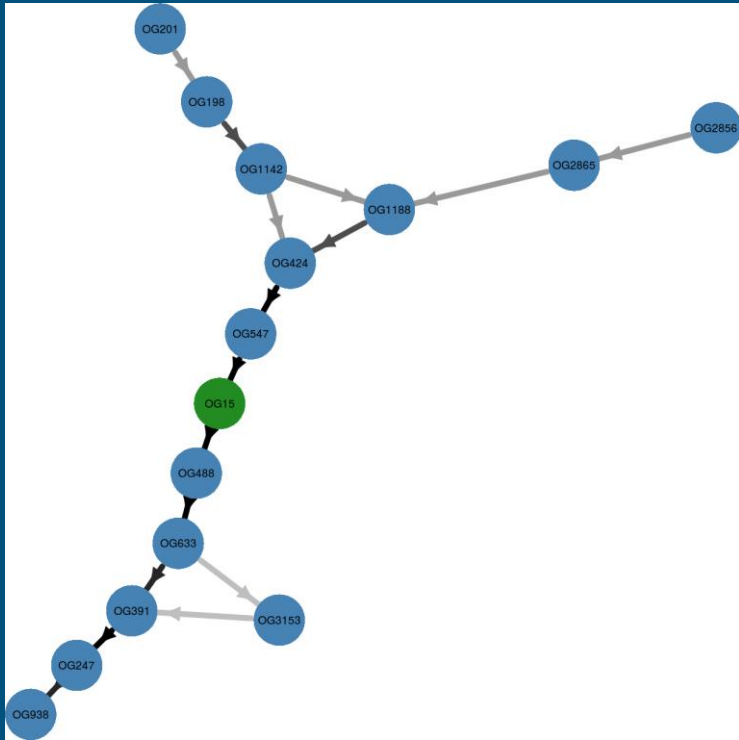
-Faster and easier way to create pangenomes for analyses

GATTCGATTAG	->	ATT: 2
		CGA: 1
		GAT: 2
		TAG: 1
		TCG: 1
		TTA: 1
		TTC: 1



# Example

-Comparison between pneumoniae and hyopneumoniae



# Summary

- Bioconductor is used for analysis, visual aid, and comprehension of genomic data
- Comparative Genomics is used to compare various species genomes to one another to look for similarities in evolution and come up with ways to fight human diseases
- FindMyFriends simplified version of BLAST in creating pangenomes. Based on decomposition of sequences into K-mer vectors. Looks at similarity, neighborhood similarity, and sequence length in comparing.
- Greatly used by researchers and evolutionists to identify how animals may have looked, behaved and their biology

# References

---

## Pictures: Google Images

-Pederson, Thomas Lin. "Creating Pangenomes Using FindMyFriends." Bioconductor, Bioconductor, 30 Oct. 2017, [www.bioconductor.org/packages/devel/bioc/vignettes/FindMyFriends/inst/doc/FindMyFriends\\_intro.html](http://www.bioconductor.org/packages/devel/bioc/vignettes/FindMyFriends/inst/doc/FindMyFriends_intro.html).

-"Comparative Genomics." National Human Genome Research Institute, NIH, 3 Nov. 2015, [www.genome.gov/11509542/comparative-genomics-fact-sheet/](http://www.genome.gov/11509542/comparative-genomics-fact-sheet/).

-U.S. Nation Library of Medicine. "What Is a Genome?" Genetics Home Reference, 14 Nov. 2017, [ghr.nlm.nih.gov/primer/hgp/genome](http://ghr.nlm.nih.gov/primer/hgp/genome).

-