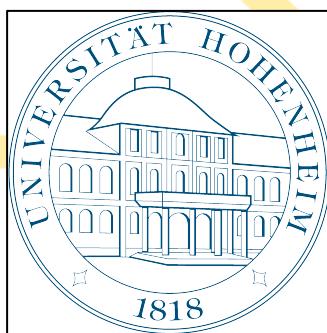
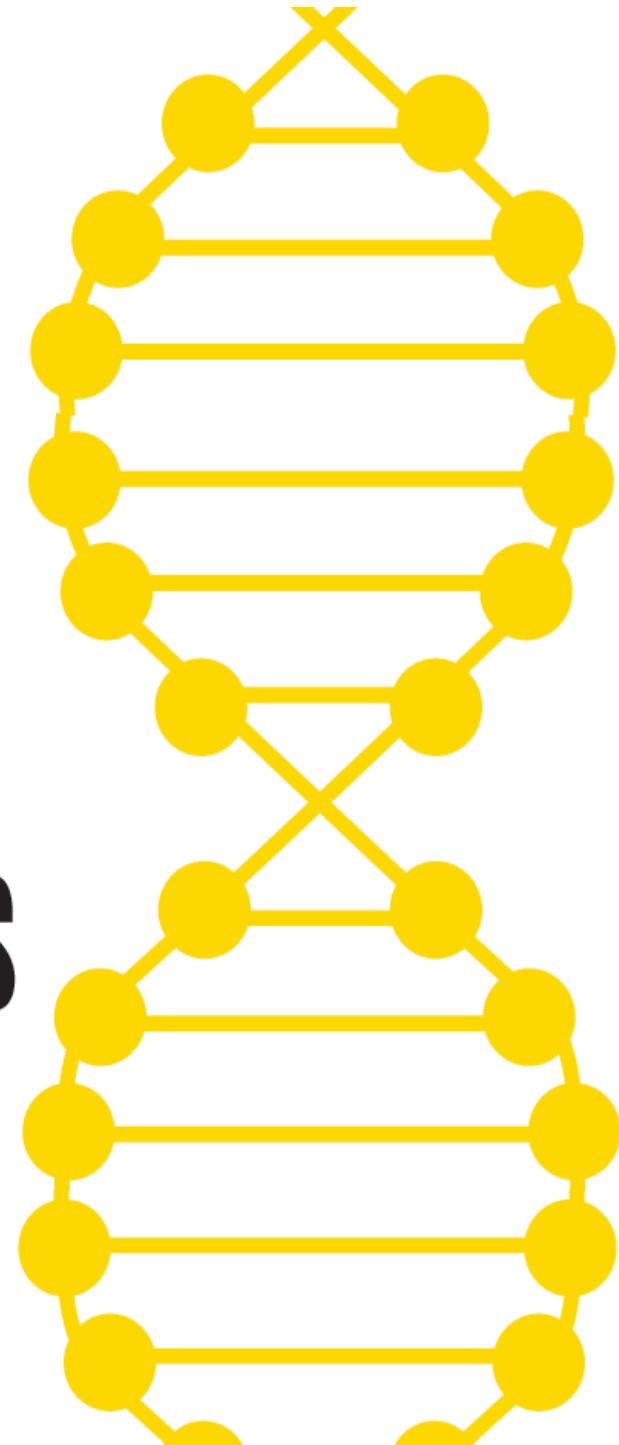


016-WS

Do-it-yourself Microbial Genome Sequence Analysis

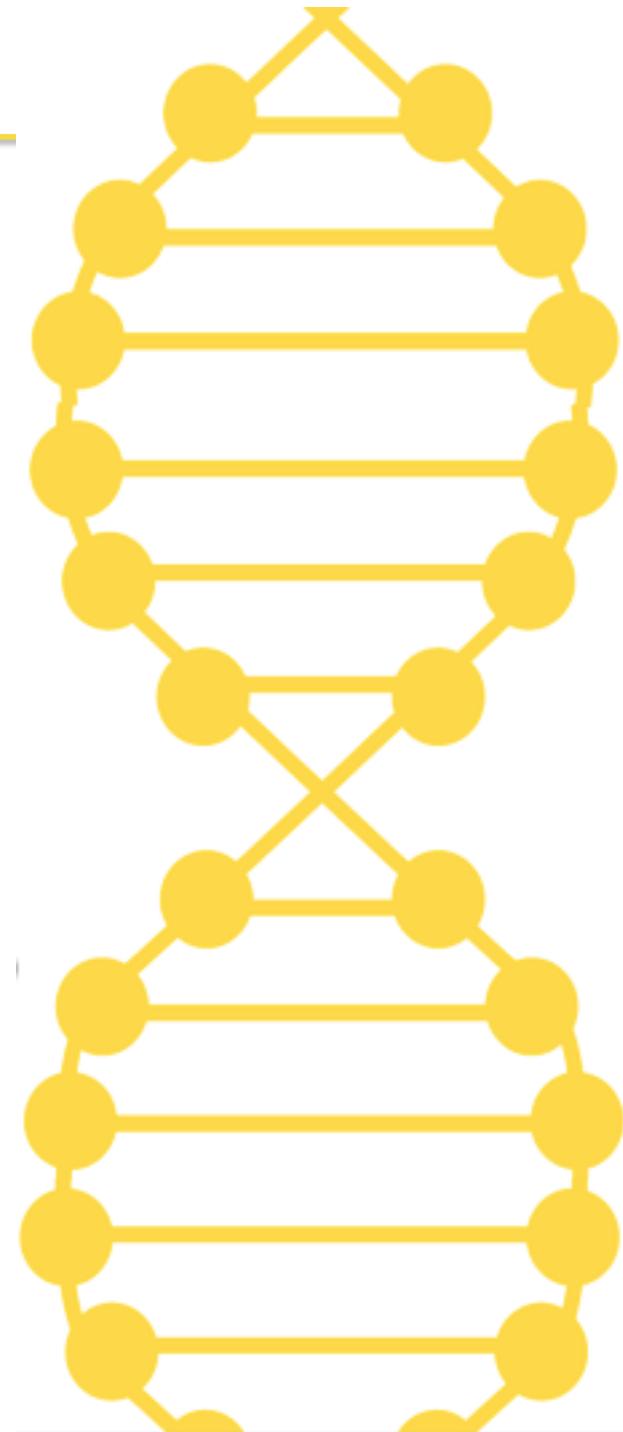


UNIVERSITY *of* MARYLAND
SCHOOL OF MEDICINE
INSTITUTE FOR GENOME SCIENCES



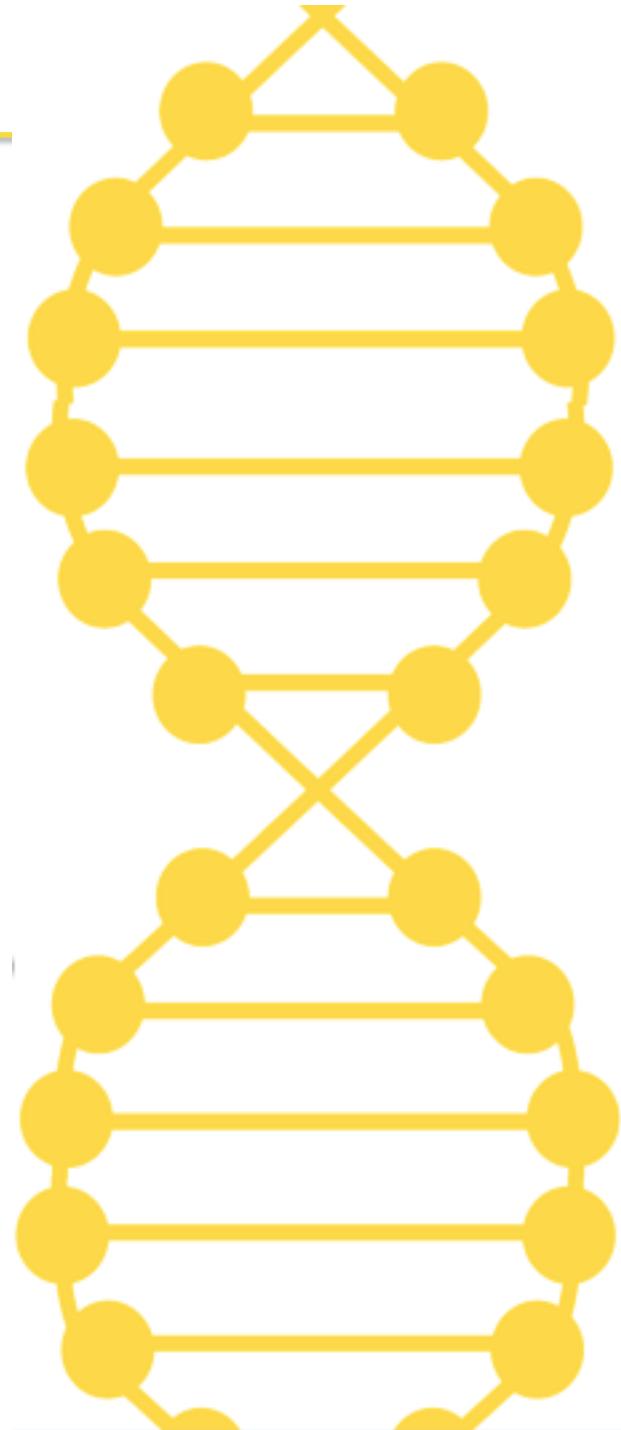
Workshop people

- **Co-conveners/presenters:**
 - W. Florian Fricke
 - Michelle Gwinn Giglio
- **Presenters**
 - Cesar Arze
 - Hervé Tettelin
 - Tracy Hazen
- **Instructional Assistants**
 - Shaun Adkins
 - Suvarna Nadendla



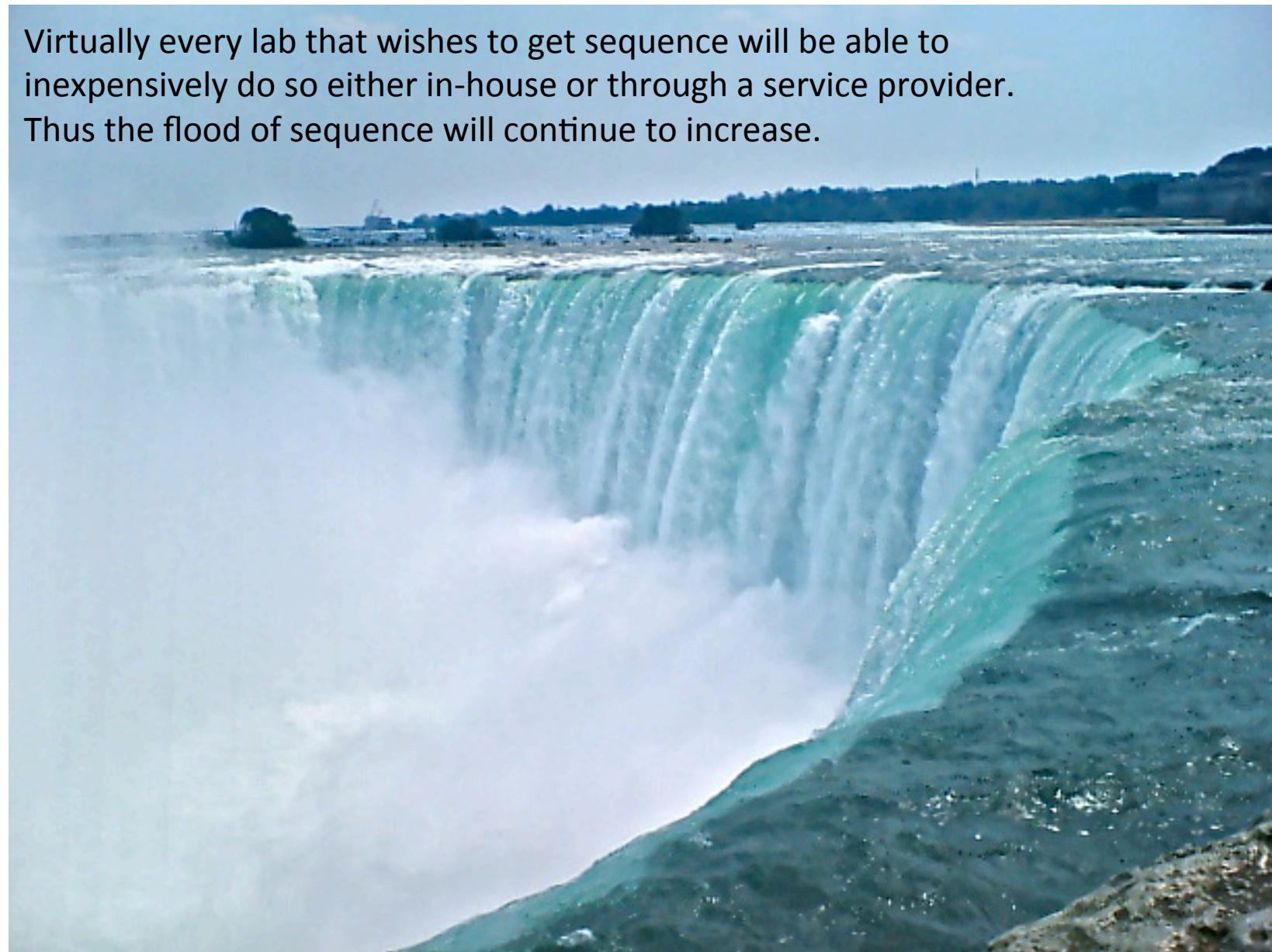
What we'll do today:

- Review two resources for analysis of prokaryotic sequences:
 - Analysis Engine
 - CloVR
- Describe two pipelines and associated visualization tools:
 - Genome annotation pipeline and Manatee
 - Comparative genomics pipeline and Sybil
- Run pipelines with CloVR
- Present research that has made use of these resources



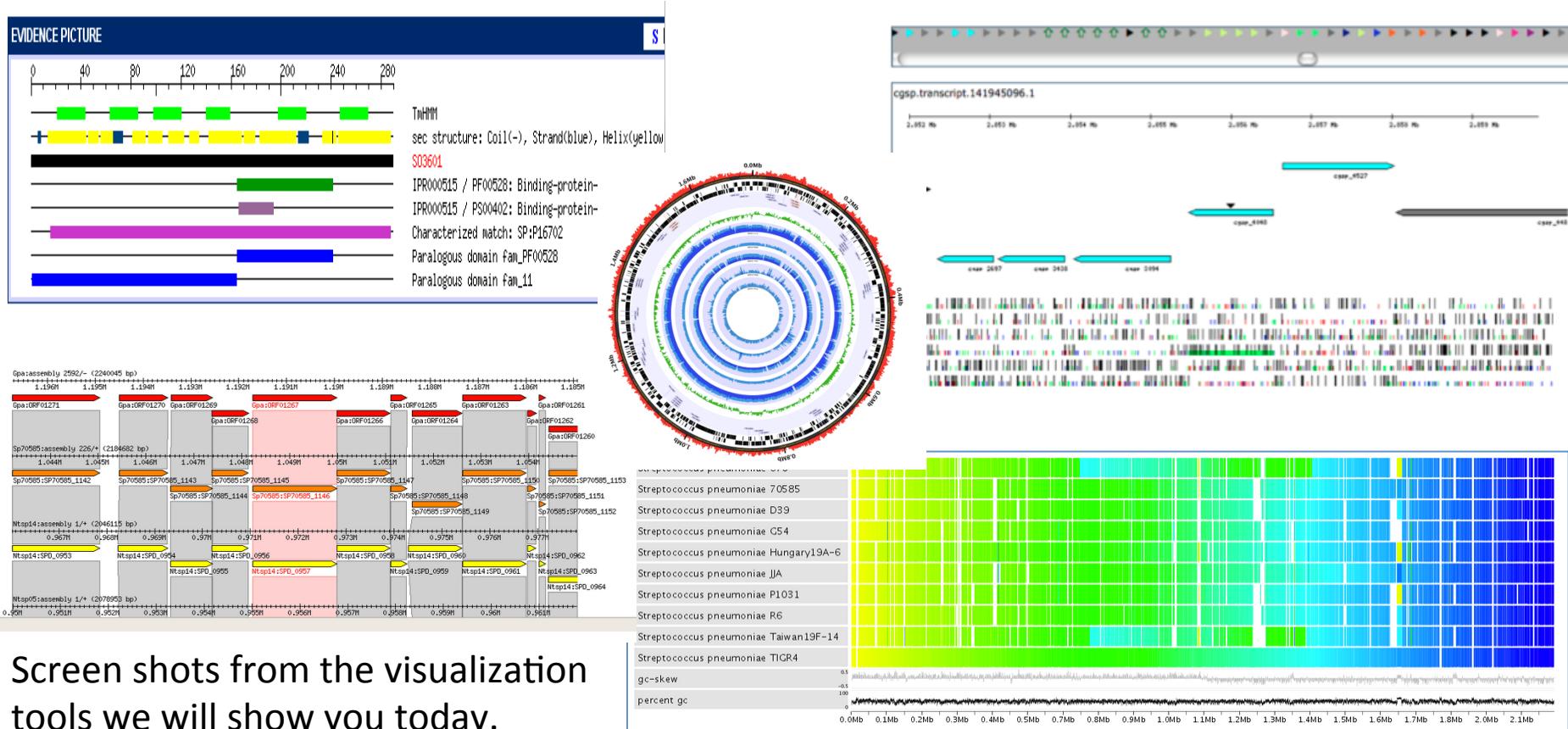
Generating sequence is now easy and inexpensive

Virtually every lab that wishes to get sequence will be able to inexpensively do so either in-house or through a service provider. Thus the flood of sequence will continue to increase.



The Analysis Challenge

- How can we provide users with the analysis tools they need?
- How do we provide users with computational resources to handle large datasets?



Screen shots from the visualization tools we will show you today.

Ways we can answer that challenge:

- Provide a service in which we run our pipelines for users on user-submitted data
- Provide remote access to our pipelines so that users can run them on cloud-based computational infrastructures
- Provide downloadable versions of the pipelines so that users can have complete control of the process.

- Our goal is to do all three!
 - We've reached that goal for the two pipelines we will describe today
(and other pipelines are available with one or two options as we continue towards the goal of all 3 options for all)

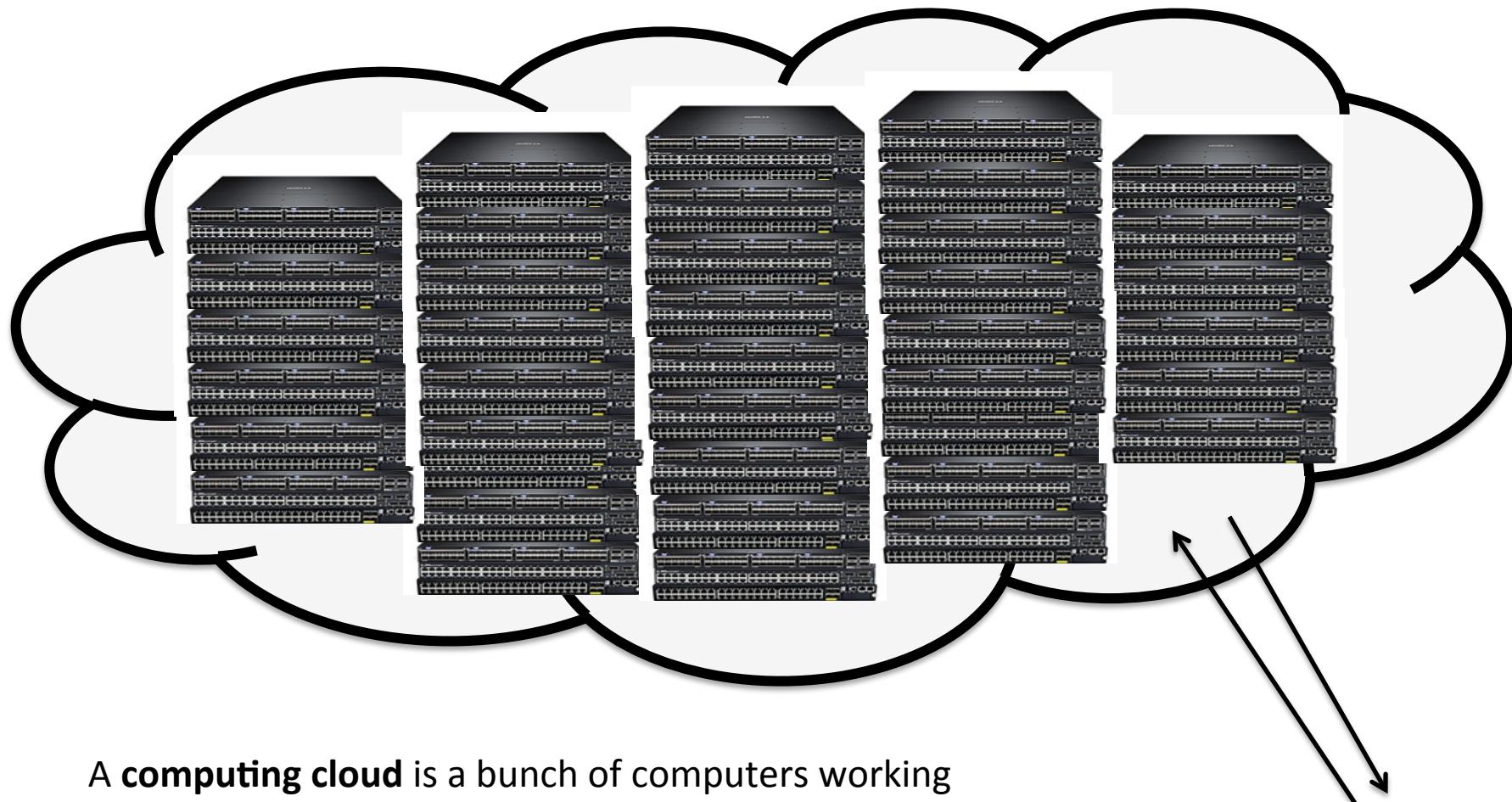
The Trans-Alaska oil pipeline is supported on sliders where it crosses the Denali Fault line, allowing it to shift as needed.



Analysis Engine plus CloVR

- Two projects working synergistically to provide these options to users.
- Analysis Engine
 - Analysis-as-a-service
 - Pipelines are run by IGS staff
 - Pipelines are also accessible on remote computing resources
 - Pipelines are run by users but within a controlled environment on IGS computing infrastructure and soon on Amazon EC2
 - Visualization tools can be
 - Accessed remotely
 - Downloaded locally
- CloVR
 - Provides pipelines in a completely portable and distributable form
 - Can be installed locally or on a cloud
 - Increases flexibility and control for the users

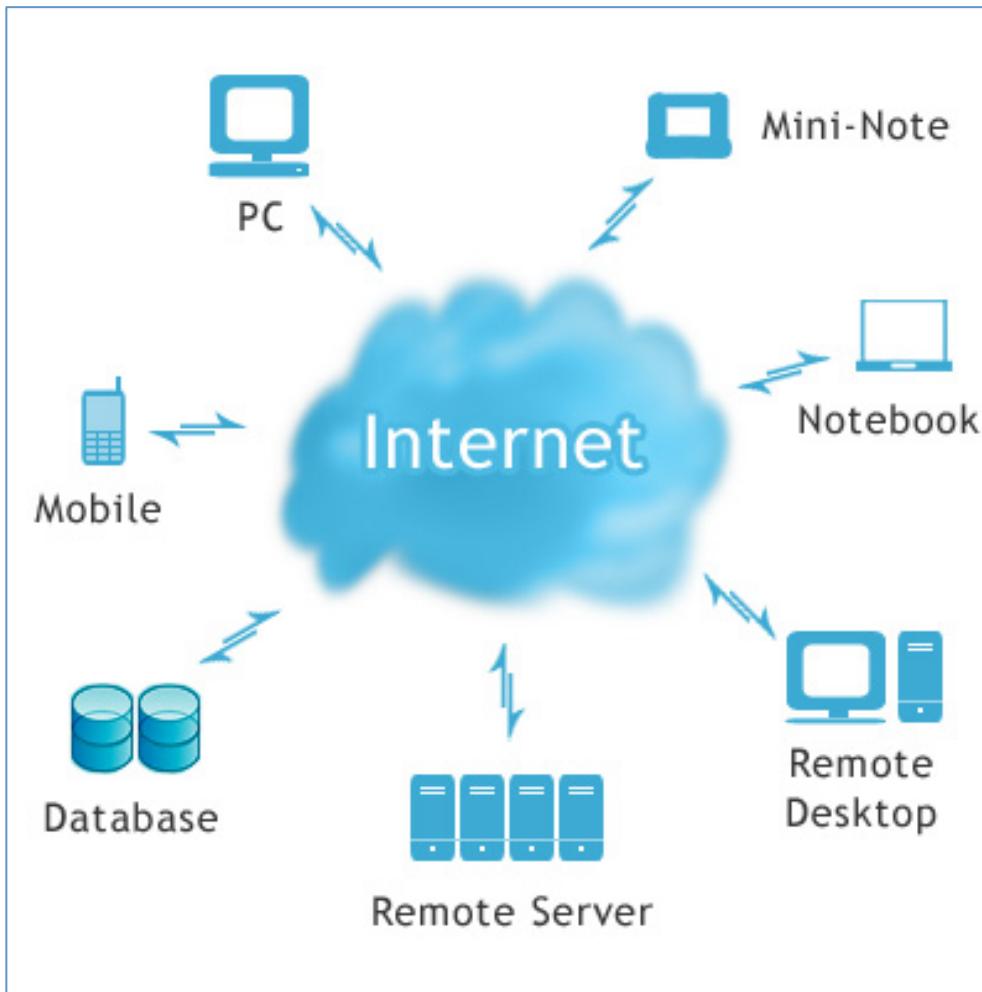
A Computing Cloud



A **computing cloud** is a bunch of computers working together, generally located remotely from users, where users can sign up to use a portion of the cpu power and run either prepackaged or custom software. Interactions between the user and cloud are via the internet.



A Computing Cloud



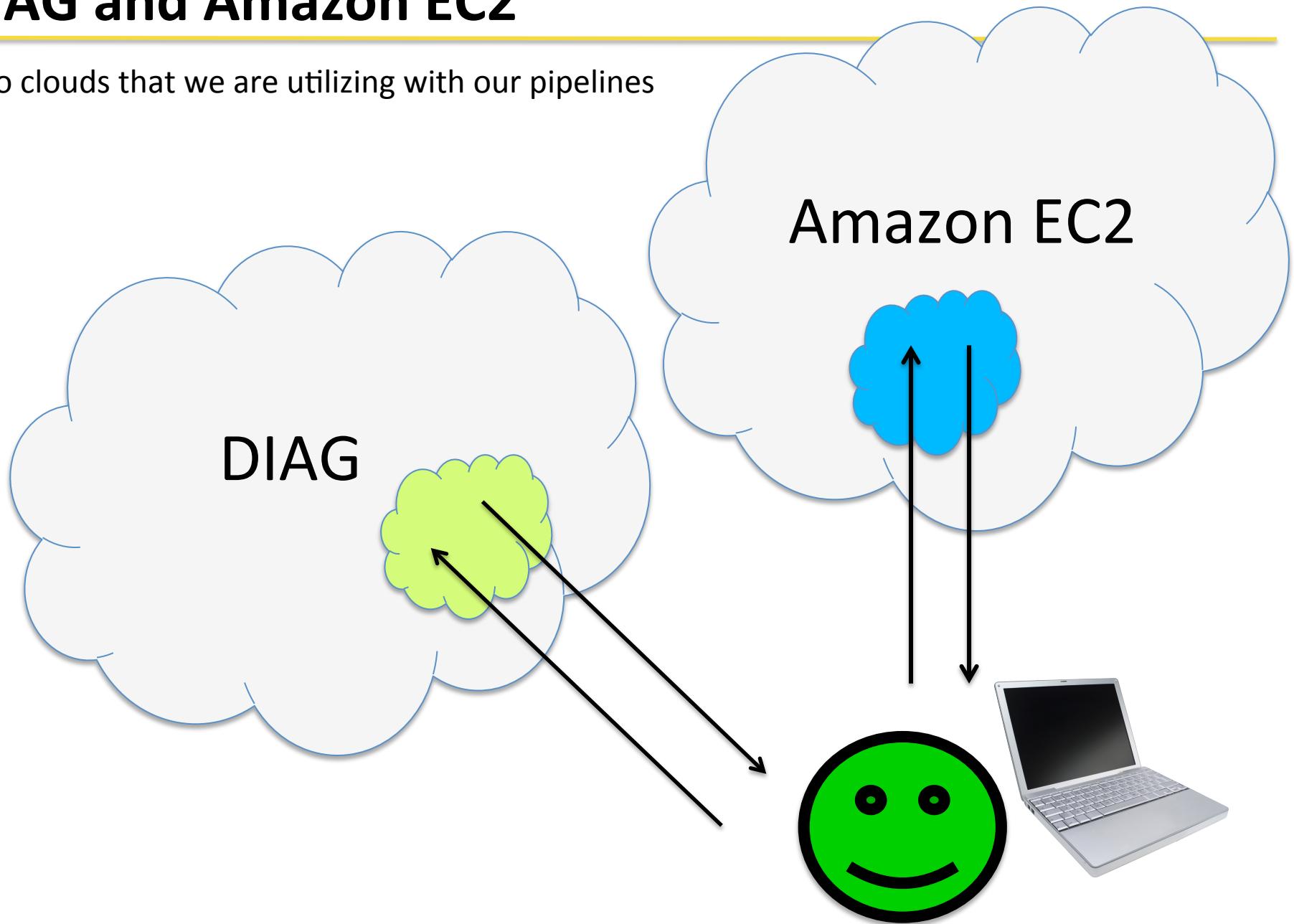
If you use any of these...



...you are using a
computing Cloud

DIAG and Amazon EC2

Two clouds that we are utilizing with our pipelines



EC2



Amazon Elastic Compute Cloud (Amazon EC2)

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers.

Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use. Amazon EC2 provides developers the tools to build failure resilient applications and isolate themselves from common failure scenarios.

DIAG: Data Intensive Academic Grid

PIs: Owen White, Anup Mahurkar, <http://diagcomputing.org>

DIAG is a cloud computing infrastructure with multiple access methods utilizing open source software for use in the biosciences field. It is free to academic researchers. It was originally funded by NSF.



We will be using
DIAG later today
in the workshop.

DIAG: Data Intensive Academic Grid

PIs: Owen White, Anup Mahurkar, <http://diagcomputing.org>

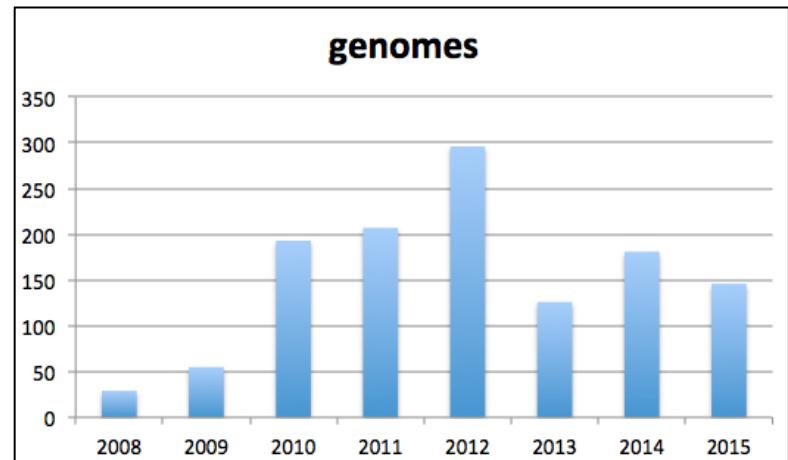
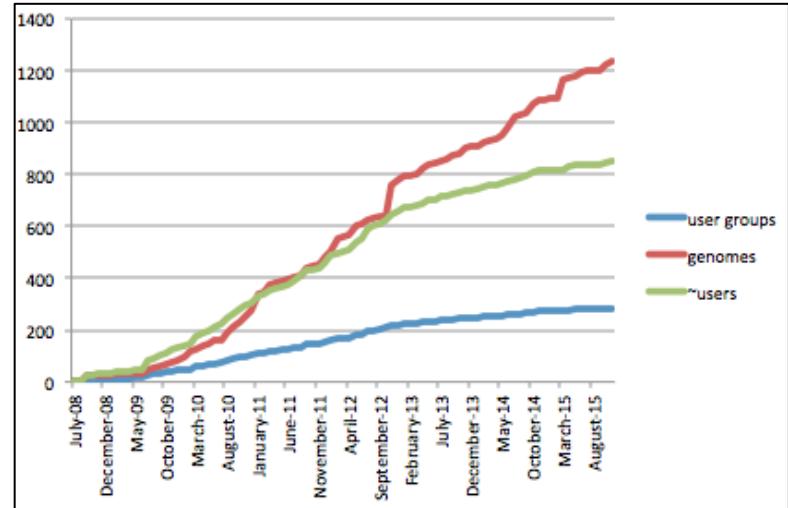
DIAG is a cloud computing infrastructure with multiple access methods utilizing open source software for use in the biosciences field. It is free to academic researchers. It was originally funded by NSF.



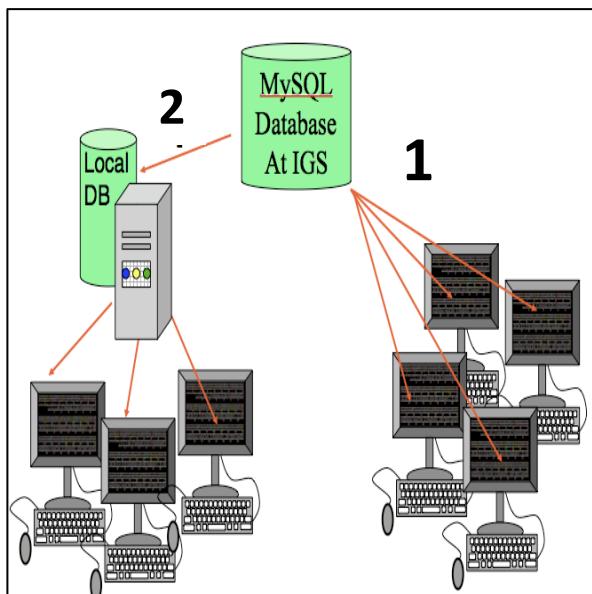
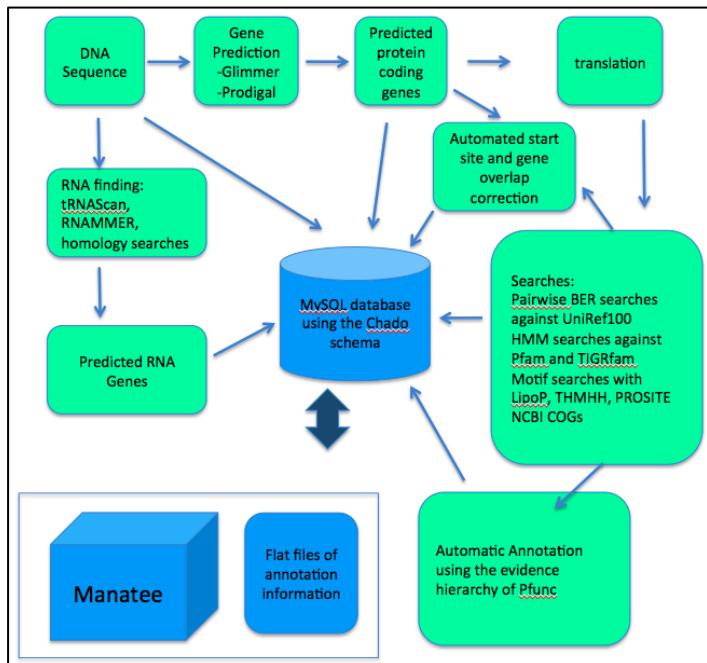
**We will be using
DIAG later today
in the workshop.**

Analysis Engine: very brief history

- Project started at IGS in 2008 as the ‘Annotation Engine’
 - prokaryotic single genome annotation
 - Manatee visualization tool
 - (Grant was transferred from TIGR)
- Renewal funding in 2013
 - Aim 1: Expansion of services, make users more independent
 - Aim 2: Third party tool integration
 - Aim 3: Outreach/training/education



AE Signature Service #1: Single genome annotation



Users who opt to send us their sequence for analysis have 3 options for results:

- Option 1**

Your database resides at IGS. We provide you a password-protected account to Manatee installed at IGS.

(By far the most popular option.)

- Option 2**

We place a MySQL version of your database and files onto an ftp site. You download it and the Manatee VM for local installation

- Option 3**

File downloads:

gff3, gbk, all annotation tab-delimited, multifasta protein/nucleotide, Gene Ontology gene association, more...

Manatee (hands-on later)

GENE CURATION INFORMATION

ORF04813 (SO2740)

- > View BER Searches
- asmb1_id: 7974
- > Reload Page

Select Display | Select Function

HMM

TIGR00433: biotin synthase gene_sym: bioB ec#: 2.8

Isology: equivalent

Trusted cutoff: 300.00 Gathering cutoff: 300.00 Noise cutoff: 50.00

Trusted cutoff2: 300.00 Gathering cutoff2: 300.00 Noise cutoff2: 50.00

View Alignment Coords HMM Coords Score Expect Curation

> align page 18-313

GENE IDENTIFICATION

gene name: biotin synthase

gene_sym: bioB

> GO:0004076 add biotin synthase activity (F)

> GO:0009102 add biotin biosynthesis (proc)

Biotin		Role id: 77							
C	seq id	gene_id	locus	end5	end3	gene name	gene symbol	ec	other
cgsa.assembly.1	cgsp_4048		2856763	2855711		biotin synthase	bioB	2.8.1.6	
cgsa.assembly.1	cgsp_4527		2856886	2858271		adenosylmethionine-8-amino-7-oxononanoate transaminase	bioA	2.6.1.62	
cgsa.assembly.1	cgsp_2852		4821460	4822251		putative pimeloyl-CoA transferase BioH	bioH		
cgsa.assembly.1	cgsp_2697		2853261	2852986		dethiobiotin synthase	bioD	6.3.3.3	

Biosynthesis of cofactors, prosthetic groups, and carriers

Folic acid		Role id: 78								
C	seq id	gene_id	locus	end5	end3	gene name	gene symbol	ec	other roles	start_edit
cgsa.assembly.1	cgsp_1064		1241974	1242807		dihydropteroate synthase	folP	2.5.1.15		
cgsa.assembly.1	cgsp_1480		901508	901020		2-amino-4-hydroxy-6-hydroxymethylidihydropteridine Pyrophosphokinase	folK	2.7.6.3		
cgsa.assembly.1	cgsp_3336		1342213	1342563		dihydroneopterin aldolase	folB	4.1.2.25		
cgsa.assembly.1	cgsp_4383		1342732	1343109		2-amino-4-hydroxy-6-hydroxymethylidihydropteridine Pyrophosphokinase	folK	2.7.6.3		
cgsa.assembly.1	cgsp_4558		2336915	2335512		aminodeoxychorismate synthase, component I	pabB	6.3.5.8		
cgsa.assembly.1	cgsp_1154		4430777	4431427		GTP cyclohydrolase I	folE	3.5.4.16		
cgsa.assembly.1	cgsp_3622		2752170	2751361		aminodeoxychorismate lyase	pabC	4.1.3.38		

Biosynthesis of cofactors, prosthetic groups, and carriers

Heme, porphyrin, and cobalamin		Role id: 79								
C	seq id	gene_id	locus	end5	end3	gene name	gene symbol	ec	other	
cgsa.assembly.1	cgsp_3341		4808057	4808959		protoheme IX farnesytransferase	cycE	2.5.1.-		
cgsa.assembly.1	cgsp_3703		1078726	1079349		cob(I)lyric acid a,c-diamide adenosyltransferase	cobO	2.5.1.17		
cgsa.assembly.1	cgsp_4255		3984756	3983506		glutamyl-tRNA reductase	hemA			
cgsa.assembly.1	cgsp_3706		4504487	460551		uracilnucleoside decarboxylase	hemF	4.1.1.37		

BER SKIM

Bolu | View BER Searches | search date: Wed Oct 23 12:59:20 2002 | Refresh Searches

accession	%sim	length	description	p-value
OMNI:SO2740	100.0	349	biotin synthase (SheWella oneidensis MR-1)	1.5e-176
SP:P36569	80.7	340	Biotin synthase (EC 2.3.1.6)(Biotin synthetase), (Serratia	2.5e-119
SP:PI2996	79.7	342	Biotin synthase (EC 2.3.1.6)(Biotin synthetase), (Escherich	7.2e-120
GP:145425	79.7	342	biotin synthetase (Escherichia coli)	1.5e-119
GP:12620127	79.4	342	biotin synthase BioB (uncultured bacterium pCosHE2)	1.5e-119
OMNI:NTL03EC0855	79.4	342	biotin synthetase (Escherichia coli O157:H7 VT2 Sakai) [OGP 13	5.1e-119
OMNI:NTL01YP1094	81.0	340	biotin synthetase (Yersinia pestis CO92)[OMNI:NTL02P2P986 biot	8.3e-119
GP:12620099	79.5	340	BioB-like protein (uncultured bacterium pCosFS1)	9.5e-118
OMNI:NTL02EC0848	79.1	342	biotin synthetase, sulfur insertion? (Escherichia coli O157:H	2.2e-118
SP:Q47867	70.7	330	Biotin synthase (EC 2.3.1.6) Biotin synthetase), (Erwinia h	3.6e-118
			[16] (Biotin synthetase), (Salmonell	5.1e-119
			oleare El Tor N16961)[OGP 9655583lg	5.1e-119
			[17] iella enterica serovar Typhi CT18)[OG	1.1e-118
			[18] nas aeruginosa PAO1)[CGP 89946364lgbl	7.7e-116
			[19] (uncultured bacterium pCosAS1)	9.1e-113
			[20] onas campestris pv. campestris ATCC3	2.8e-111
			[21] onas axonopodis pv. citri 306)[OGP 21	6.6e-110
			[22] 185lgblAAM6	1.4e-109
			\$ [34]gblAAF8	8.4e-110
			[23] planaceum GMI	4.7e-109
			Buchnera	1.1e-107
			[24] us)	1.6e-106
			[25] 1342521lg	3.0e-105
			S03601	6.3e-105
			IPR000515 / PF00528: Binding-protein-dependent tra	
			IPR000515 / PS00402: Binding-protein-dependent tra	
			Characterized match: SP:PI6702	
			Paralogous domain fam_PF00528	
			Paralogous domain fam_11	

EVIDENCE PICTURE

function

add go id ev code

ISS ISS ISS ISS ISS ISS

0 40 80 120 160 200 240 280

TnHMM sec structure: Coil(-), Strand(blue), Helix(yellow) 984969/gblAAL

IPR000515 / PF00528: Binding-protein-dependent tra

IPR000515 / PS00402: Binding-protein-dependent tra

Characterized match: SP:PI6702

Paralogous domain fam_PF00528

Paralogous domain fam_11

cgsp.transcript.141945096.1

2.952 Mb 2.953 Mb 2.954 Mb 2.955 Mb 2.956 Mb 2.957 Mb 2.958 Mb

cgsp_3341

cgsp_3703

cgsp_4255

cgsp_3706

cgsp_1154

cgsp_1480

cgsp_2852

cgsp_3336

cgsp_4527

cgsp_4558

cgsp_1064

cgsp_3622

cgsp_2697

cgsp_3400

cgsp_3494

cgsp_3496

cgsp_3498

cgsp_3500

cgsp_3502

cgsp_3504

cgsp_3506

cgsp_3508

cgsp_3510

cgsp_3512

cgsp_3514

cgsp_3516

cgsp_3518

cgsp_3520

cgsp_3522

cgsp_3524

cgsp_3526

cgsp_3528

cgsp_3530

cgsp_3532

cgsp_3534

cgsp_3536

cgsp_3538

cgsp_3540

cgsp_3542

cgsp_3544

cgsp_3546

cgsp_3548

cgsp_3550

cgsp_3552

cgsp_3554

cgsp_3556

cgsp_3558

cgsp_3560

cgsp_3562

cgsp_3564

cgsp_3566

cgsp_3568

cgsp_3570

cgsp_3572

cgsp_3574

cgsp_3576

cgsp_3578

cgsp_3580

cgsp_3582

cgsp_3584

cgsp_3586

cgsp_3588

cgsp_3590

cgsp_3592

cgsp_3594

cgsp_3596

cgsp_3598

cgsp_3600

cgsp_3602

cgsp_3604

cgsp_3606

cgsp_3608

cgsp_3610

cgsp_3612

cgsp_3614

cgsp_3616

cgsp_3618

cgsp_3620

cgsp_3622

cgsp_3624

cgsp_3626

cgsp_3628

cgsp_3630

cgsp_3632

cgsp_3634

cgsp_3636

cgsp_3638

cgsp_3640

cgsp_3642

cgsp_3644

cgsp_3646

cgsp_3648

cgsp_3650

cgsp_3652

cgsp_3654

cgsp_3656

cgsp_3658

cgsp_3660

cgsp_3662

cgsp_3664

cgsp_3666

cgsp_3668

cgsp_3670

cgsp_3672

cgsp_3674

cgsp_3676

cgsp_3678

cgsp_3680

cgsp_3682

cgsp_3684

cgsp_3686

cgsp_3688

cgsp_3690

cgsp_3692

cgsp_3694

cgsp_3696

cgsp_3698

cgsp_3700

cgsp_3702

cgsp_3704

cgsp_3706

cgsp_3708

cgsp_3710

cgsp_3712

cgsp_3714

cgsp_3716

cgsp_3718

cgsp_3720

cgsp_3722

cgsp_3724

cgsp_3726

cgsp_3728

cgsp_3730

cgsp_3732

cgsp_3734

cgsp_3736

cgsp_3738

cgsp_3740

cgsp_3742

cgsp_3744

cgsp_3746

cgsp_3748

cgsp_3750

cgsp_3752

cgsp_3754

cgsp_3756

cgsp_3758

cgsp_3760

cgsp_3762

cgsp_3764

cgsp_3766

cgsp_3768

cgsp_3770

cgsp_3772

cgsp_3774

cgsp_3776

cgsp_3778

cgsp_3780

cgsp_3782

cgsp_3784

cgsp_3786

cgsp_3788

cgsp_3790

cgsp_3792

cgsp_3794

cgsp_3796

cgsp_3798

cgsp_3800

cgsp_3802

cgsp_3804

cgsp_3806

cgsp_3808

cgsp_3810

cgsp_3812

cgsp_3814

cgsp_3816

cgsp_3818

cgsp_3820

cgsp_3822

cgsp_3824

cgsp_3826

cgsp_3828

cgsp_3830

cgsp_3832

cgsp_3834

cgsp_3836

cgsp_3838

cgsp_3840

cgsp_3842

cgsp_3844

cgsp_3846

cgsp_3848

cgsp_3850

cgsp_3852

cgsp_3854

cgsp_3856

cgsp_3858

cgsp_3860

cgsp_3862

cgsp_3864

cgsp_3866

cgsp_3868

cgsp_3870

cgsp_3872

cgsp_3874

cgsp_3876

cgsp_3878

cgsp_3880

cgsp_3882

cgsp_3884

cgsp_3886

cgsp_3888

cgsp_3890

cgsp_3892

cgsp_3894

cgsp_3896

cgsp_3898

cgsp_3900

cgsp_3902

cgsp_3904

cgsp_3906

cgsp_3908

cgsp_3910

cgsp_3912

cgsp_3914

cgsp_3916

cgsp_3918

cgsp_3920

cgsp_3922

cgsp_3924

cgsp_3926

cgsp_3928

cgsp_3930

cgsp_3932

cgsp_3934

cgsp_3936

cgsp_3938

cgsp_3940

cgsp_3942

cgsp_3944

cgsp_3946

cgsp_3948

cgsp_3950

cgsp_3952

cgsp_3954

cgsp_3956

cgsp_3958

cgsp_3960

cgsp_3962

cgsp_3964

cgsp_3966

cgsp_3968

cgsp_3970

cgsp_3972

cgsp_3974

cgsp_3976

cgsp_3978

cgsp_3980

cgsp_3982

cgsp_3984

cgsp_3986

cgsp_3988

cgsp_3990

cgsp_3992

cgsp_3994

cgsp_3996

cgsp_3998

cgsp_4000

cgsp_4002

cgsp_4004

cgsp_4006

cgsp_4008

cgsp_4010

cgsp_4012

cgsp_4014

cgsp_4016

cgsp_4018

cgsp_4020

cgsp_4022

cgsp_4024

cgsp_4026

cgsp_4028

cgsp_4030

cgsp_4032

cgsp_4034

cgsp_4036

cgsp_4038

cgsp_4040

cgsp_4042

cgsp_4044

cgsp_4046

cgsp_4048

cgsp_4050

cgsp_4052

cgsp_4054

cgsp_4056

cgsp_4058

cgsp_4060

cgsp_4062

cgsp_4064

cgsp_4066

cgsp_4068

cgsp_4070

cgsp_4072

cgsp_4074

cgsp_4076

cgsp_4078

cgsp_4080

cgsp_4082

cgsp_4084

cgsp_4086

cgsp_4088

cgsp_4090

cgsp_4092

cgsp_4094

cgsp_4096

cgsp_4098

cgsp_4100

cgsp_4102

cgsp_4104

cgsp_4106

cgsp_4108

cgsp_4110

cgsp_4112

cgsp_4114

cgsp_4116

cgsp_4118

cgsp_4120

cgsp_4122

cgsp_4124

cgsp_4126

cgsp_4128

cgsp_4130

cgsp_4132

cgsp_4134

cgsp_4136

cgsp_4138

cgsp_4140

cgsp_4142

cgsp_4144

cgsp_4146

cgsp_4148

cgsp_4150

cgsp_4152

cgsp_4154

cgsp_4156

cgsp_4158

cgsp_4160

cgsp_4162

cgsp_4164

cgsp_4166

cgsp_4168

cgsp_4170

cgsp_4172

cgsp_4174

cgsp_4176

cgsp_4178

cgsp_4180

cgsp_4182

cgsp_4184

cgsp_4186

cgsp_4188

cgsp_4190

cgsp_4192

cgsp_4194

cgsp_4196

cgsp_4198

cgsp_4200

cgsp_4202

cgsp_4204

cgsp_4206

cgsp_4208

cgsp_4210

cgsp_4212

cgsp_4214

cgsp_4216

cgsp_4218

cgsp_4220

cgsp_4222

cgsp_4224

cgsp_4226

cgsp_4228

cgsp_4230

cgsp_4232

cgsp_4234

cgsp_4236

cgsp_4238

cgsp_4240

cgsp_4242

cgsp_4244

cgsp_4246

cgsp_4248

cgsp_4250

cgsp_4252

cgsp_4254

cgsp_4256

cgsp_4258

cgsp_4260

cgsp_4262

cgsp_4264

cgsp_4266

cgsp_4268

cgsp_4270

cgsp_4272

cgsp_4274

cgsp_4276

cgsp_4278

cgsp_4280

cgsp_4282

cgsp_4284

cgsp_4286

cgsp_4288

cgsp_4290

cgsp_4292

cgsp_4294

cgsp_4296

cgsp_4298

cgsp_4300

cgsp_4302

cgsp_4304

cgsp_4306

cgsp_4308

cgsp_4310

cgsp_4312

cgsp_4314

cgsp_4316

cgsp_4318

cgsp_4320

cgsp_4322

cgsp_4324

cgsp_4326

cgsp_4328

cgsp_4330

cgsp_4332

cgsp_4334

cgsp_4336

cgsp_4338

cgsp_4340

cgsp_4342

cgsp_4344

cgsp_4346

cgsp_4348

cgsp_4350

cgsp_4352

cgsp_4354

cgsp_4356

cgsp_4358

cgsp_4360

cgsp_4362

cgsp_4364

cgsp_4366

cgsp_4368

cgsp_4370

cgsp_4372

cgsp_4374

cgsp_4376

cgsp_4378

cgsp_4380

cgsp_4382

cgsp_4384

cgsp_4386

cgsp_4388

cgsp_4390

cgsp_4392

cgsp_4394

cgsp_4396

cgsp_4398

cgsp_4400

cgsp_4402

cgsp_4404

cgsp_4406

cgsp_4408

cgsp_4410

cgsp_4412

cgsp_4414

cgsp_4416

cgsp_4418

cgsp_4420

cgsp_4422

cgsp_4424

cgsp_4426

cgsp_4428

cgsp_4430</p

Manatee (hands-on later)

GENE CURATION INFORMATION

ORF04813 (SO2740)		end5/end3: 2856763 / 2855711	...
View BER Searches asmb1_id: 7974 Reload Page		HMM	...
		gene length: 300	...
		protein length: 100	...
		molecular wt: 11000	...
<input type="button" value="Select Display"/>		<input type="button" value="Select Function"/>	...

GENE IDENTIFICATION

gene name:	biotin synthase	View Alignment	Coords				
gene_sym:	bioB	align page	18-313				
Biotin		GO:0006656					
C	seq id	gene_id	locus	end5	end3	gene name	Role
(1)	cgsp_assembly_1	cgsp_4048		2856763		dihydropteroate synthase	folP
	cgsp_assembly_1	cgsp_4527		2856886		2-amino-4-hydroxy-6-hydroxymethylidihydropteridine Pyrophosphokinase	folK
	cgsp_assembly_1	cgsp_2852		4821460		dihydronicopterin aldolase	folB
	cgsp_assembly_1	cgsp_2697		2853281		2-amino-4-hydroxy-6-hydroxymethylidihydropteridine Pyrophosphokinase	folK

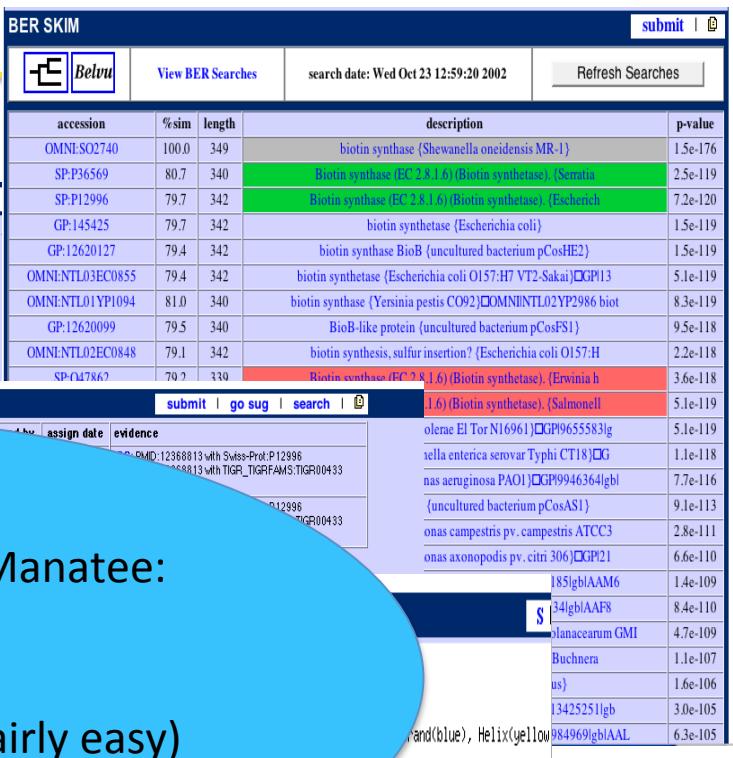
Biosynthesis of cofactors, prosthetic groups, and carriers

Folic acid		Three options					
C	seq id	gene_id	locus	end5	end3	gene name	Role
(1)	cgsp_assembly_1	cgsp_1064		1241974	1242807	dihydropteroate synthase	folP
	cgsp_assembly_1	cgsp_1480		901508	901020	2-amino-4-hydroxy-6-hydroxymethylidihydropteridine Pyrophosphokinase	folK
	cgsp_assembly_1	cgsp_3336		1342213	1342563	dihydronicopterin aldolase	folB
	cgsp_assembly_1	cgsp_4383		1342732	1343109	2-amino-4-hydroxy-6-hydroxymethylidihydropteridine Pyrophosphokinase	folK
	cgsp_assembly_1	cgsp_4558		2336915	2335512	aminodeoxychorismate synthase, component I	pabB
	cgsp_assembly_1	cgsp_1154		4430777	4431427	GTP cyclohydrolase I	folE
	cgsp_assembly_1	cgsp_3622		2752170	2751361	aminodeoxychorismate lyase	pabC

Biosynthesis of cofactors, prosthetic groups, and carriers

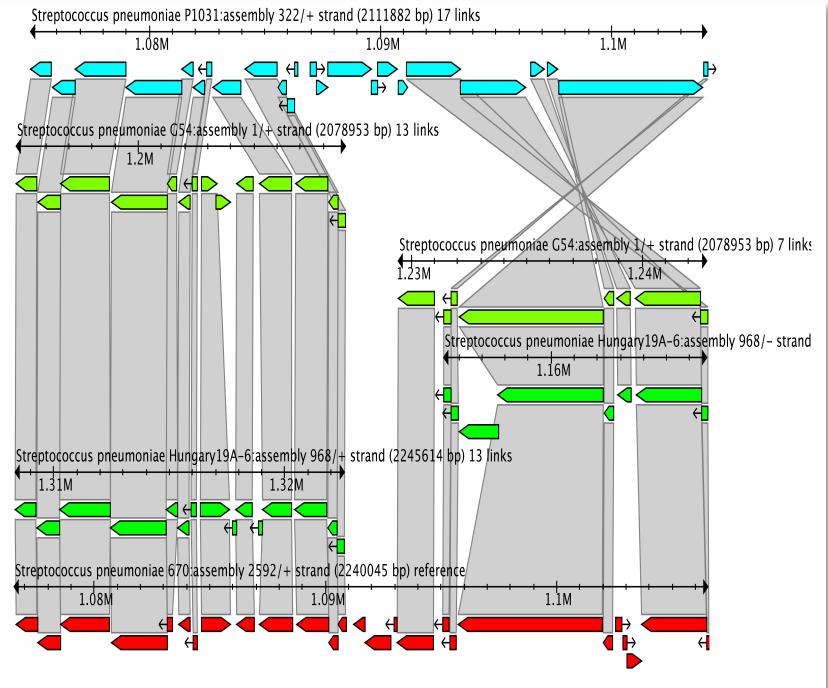
- Three options for accessing Manatee:
 - 1. Web access (super easy)
 - 2. Local installation (difficult)
 - 3. Manatee Virtual machine (fairly easy)

Heme, porphyrin, and cobalamin							Role id: 79		
C	seq id	gene_id	locus	end5	end3	gene name	gene symbol	ec	otu
	cgsp_assembly_1	cgsp_3341		4808057	4808959	protoheme IX farnesytransferase	cyoE	2.5.1.-	
	cgsp_assembly_1	cgsp_3703		1078726	1079349	cob(I)lyrinic acid a,c-diamide adenosyltransferase	cobO	2.5.1.17	
	cgsp_assembly_1	cgsp_4255		3984756	3988356	glutamyl-tRNA reductase	hemA		
	cgsp_assembly_1	cgsp_3206		450487	460651	uncharacterized deacetylase	hemF	4.1.1.37	



AE Signature Service #2: Comparative Genomics

- Mugsy
 - Whole genome nucleotide alignments
 - For genomes that are fairly closely related
 - You'll hear more about this pipeline later today
- Jaccard Clustering
 - Protein-based clusters
 - Can be used for more either closely related or divergent genomes
- Sybil visualization tool
 - You will do exercises with Sybil today

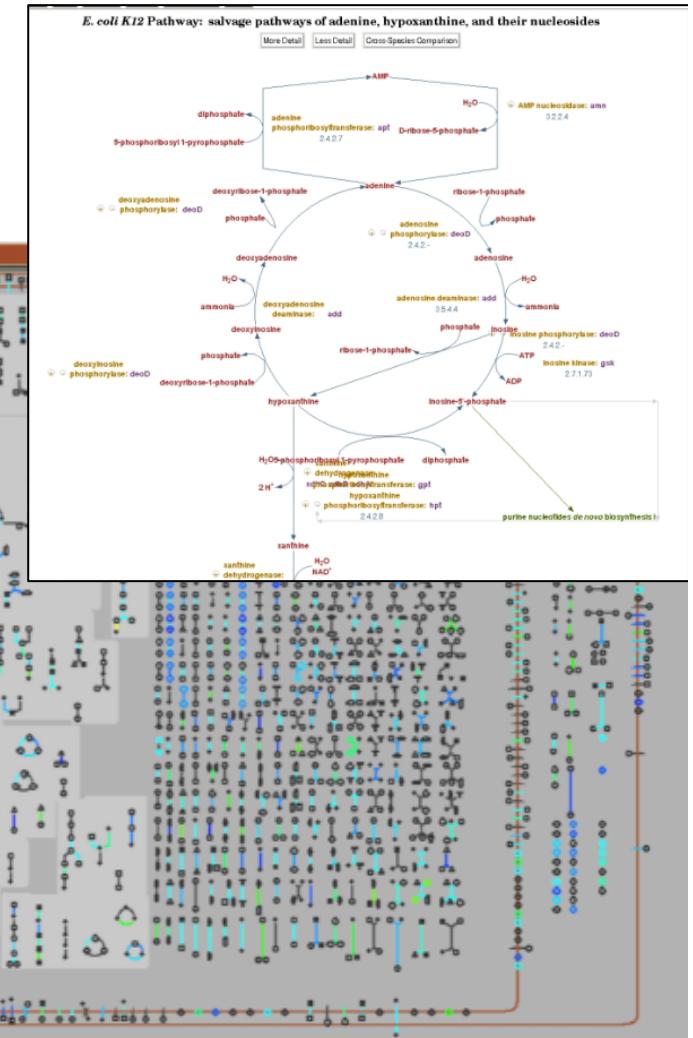
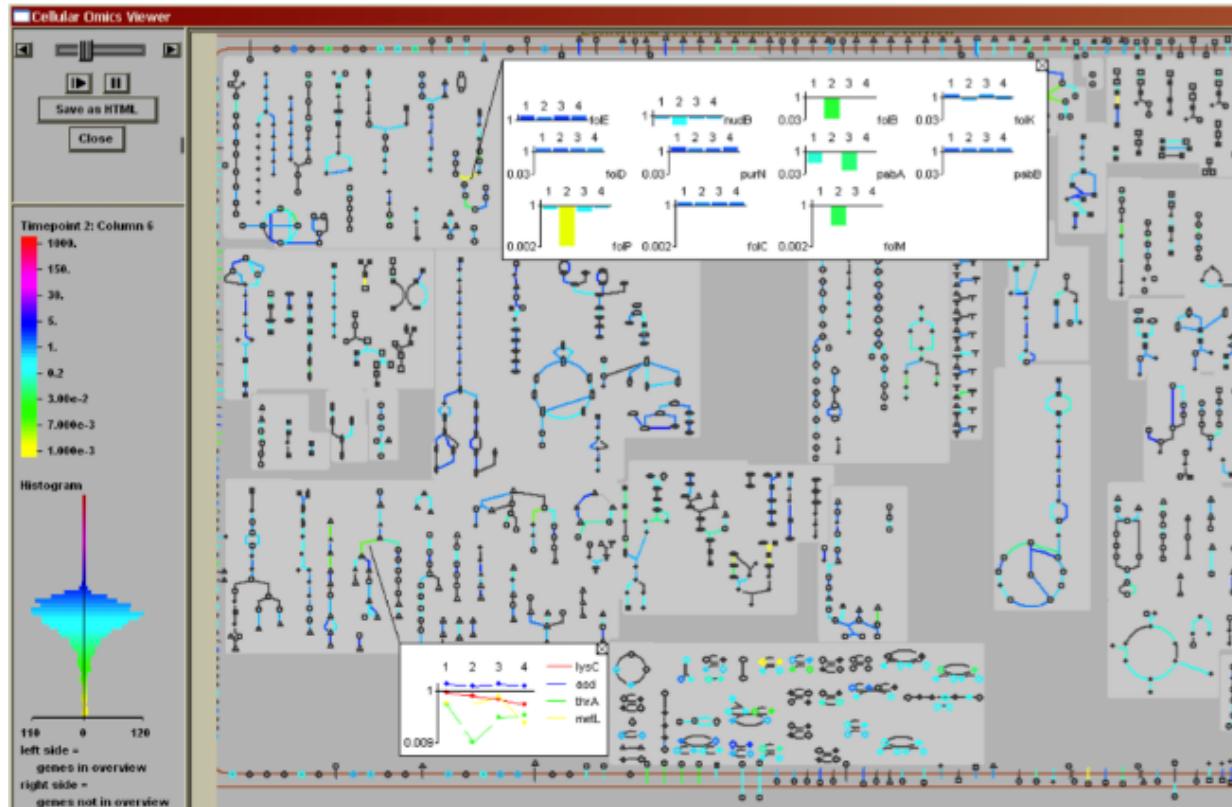


Analysis Engine

**Additional services not covered
during today's workshop...**

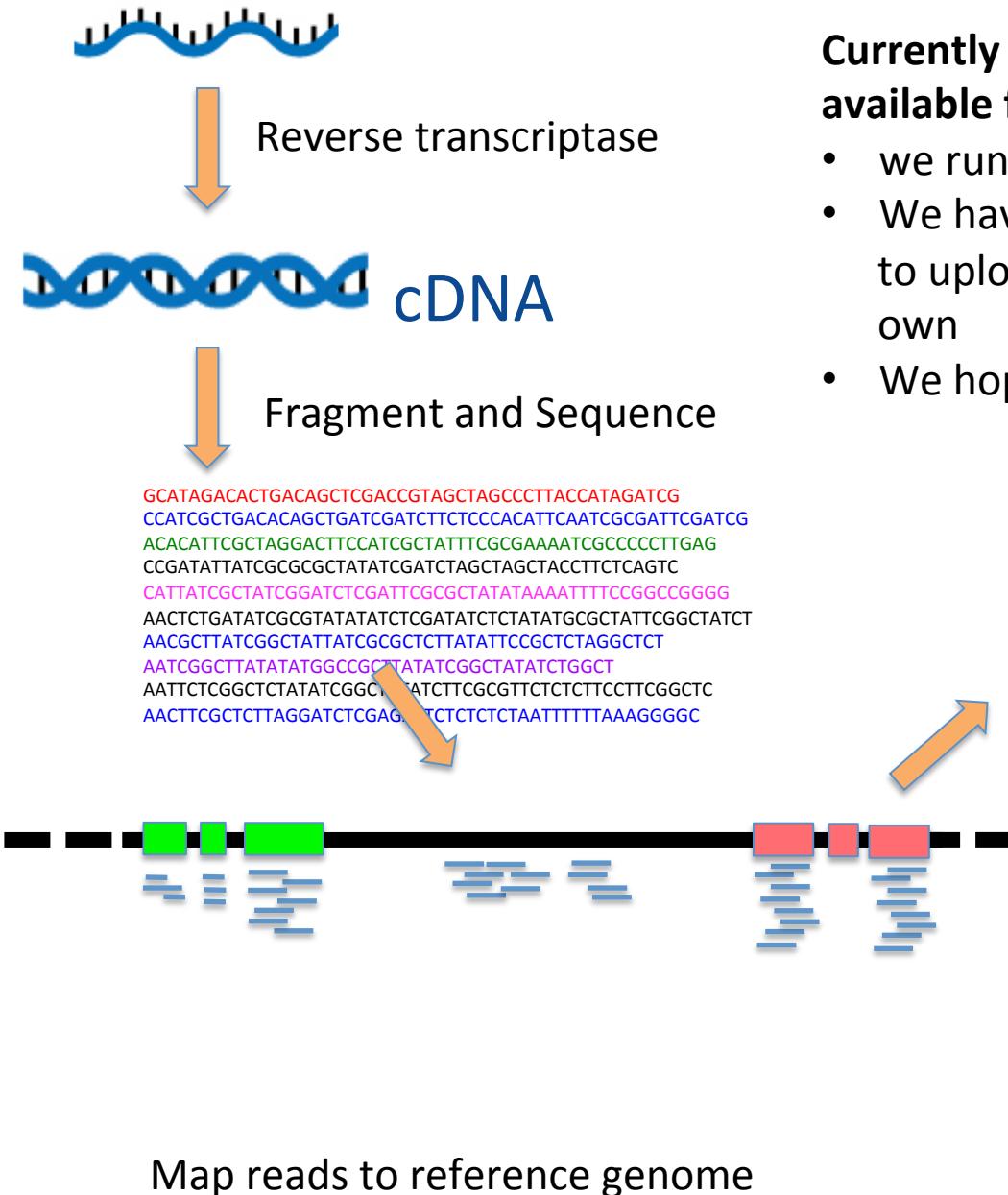
Analysis Engine – Pathway Tools

We provide Pathway Tools metabolic reconstruction analysis. We are working with Peter Karp and SRI to expand our capabilities in this area and to more fully link AE and PT.



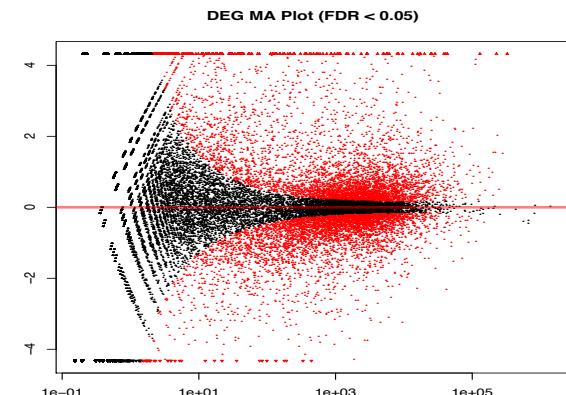
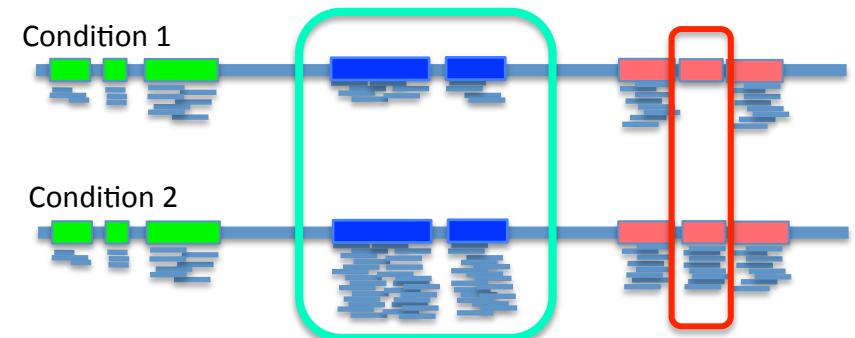
<http://bioinformatics.ai.sri.com/ptools/ptools-overview.html>

Analysis Engine – Transcriptome Analysis



Currently we have two of the three options available for transcriptome analysis

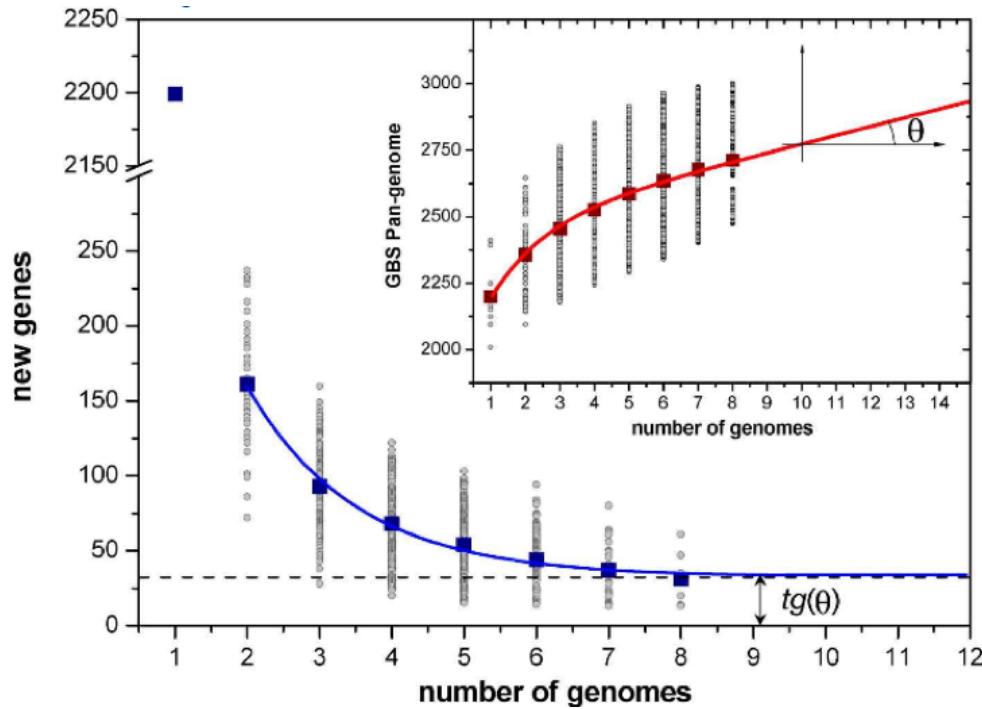
- we run the pipeline on user-submitted data
- We have a user-interface on DIAG for users to upload data and run the pipeline on their own
- We hope to have the pipeline in CloVR soon



Analysis Engine – Pangenome Analysis

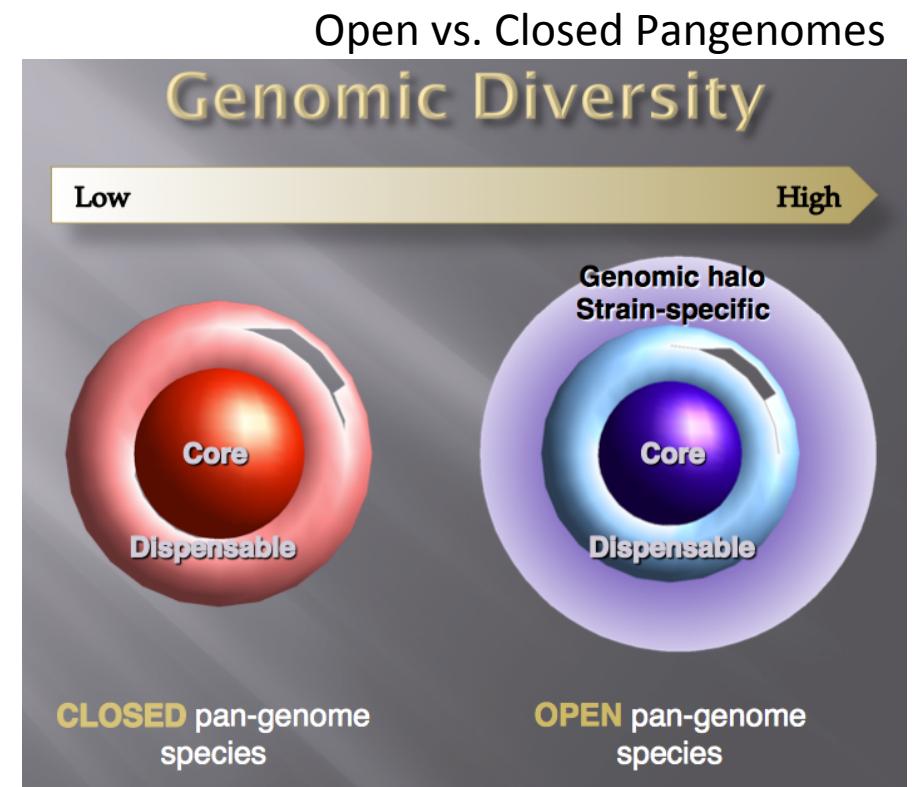
Analysis of the full complement of genes ever observed in a species – the pangenome.

Currently this pipeline is only offered as a service.



Images stolen from Hervé Tettelin

S. agalactiae (GBS) new genes and pangenome



A little advertising before I go...

Other training opportunities



Education – Training Workshops

Training Workshops Overview

The Institute for Genome Sciences offers regular training courses on 'omics technologies, bioinformatics, and programming. Registration for all 2015 workshops is now open. Please click on the titles below for more information.

IGS Introduction to Genomics and Bioinformatics

This workshop provides an introduction to the methods and tools used in genome analysis. It is designed for attendees who have fundamental knowledge of biology, but no prior genomics experience is expected or required. Topics include sequencing applications/technologies, genome annotation, comparative genomics, transcriptomics, metagenomics. Hands-on exercises using popular bioinformatics tools are included for all topics covered in the course.

IGS Introduction to Programming for Bioinformatics

This workshop provides a basic introduction to three aspects of programming as applied to bioinformatics: Perl, databases, and R. No prior programming experience is expected or required. The workshop is designed to give biologists the fundamental programming tools they need to operate in a Linux/Unix environment and to manipulate data files, engage in analysis, and effectively store and retrieve their data.

IGS Metagenome Analysis Workshop

This workshop will provide attendees with in-depth training on analysis of bacterial community sequence data, both whole metagenome shotgun and 16S. Tools for community profiling, gene clustering, and annotation will be explored. An optional 4th day is included in the workshop for those who elect to bring a dataset of their own for analysis.

IGS Transcriptome Analysis Workshop

This workshop will provide training on the tools used for analysis of RNA-Seq data. Methods covered include: mapping reads to a reference and differential expression analysis.

IGS Prokaryotic Comparative Genomics Workshop

This workshop will explore the tools used for comparisons of multiple (up to dozens) of prokaryotic sequences. Tools for pangenome analysis, whole genome alignment, ortholog clustering, SNP analysis, and visualization of data will be explored.

igs.umaryland.edu – Education tab

Please visit booth #407

- IGS is hosting an exhibitor booth
- Come to learn more about the many other resources IGS has available to the community
 - Sequencing services
 - Training workshops
 - Other analysis services (besides the ones we mentioned today)
 - microbiome data and analysis
 - eukaryotic genome annotation
 - Much more...

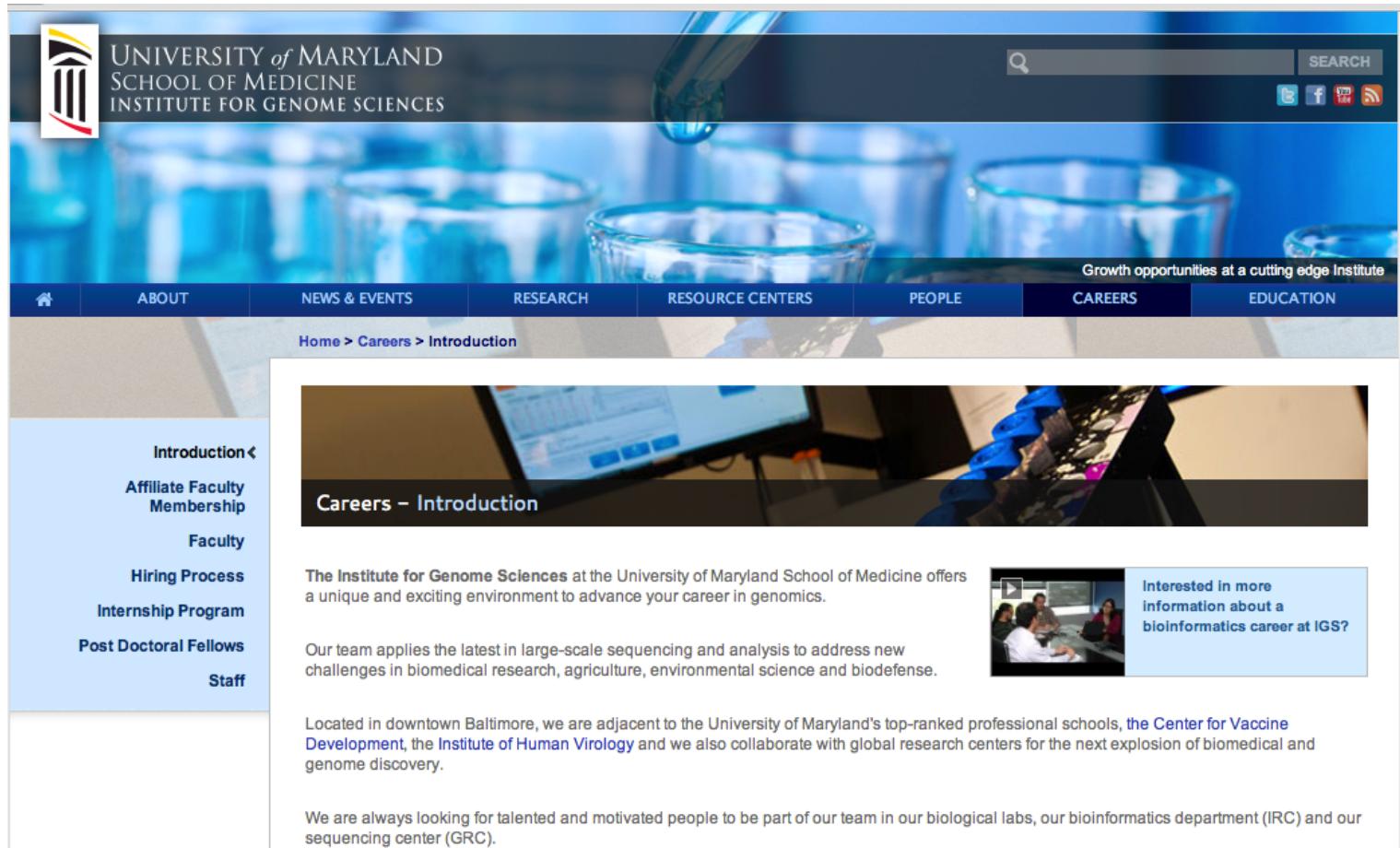


Getting organized at IGS for the booth

Employment Opportunities at IGS

- IT positions
- Software engineers
- Analysts

Please visit:
<http://www.igs.umaryland.edu/careers/>



The screenshot shows the homepage of the University of Maryland Institute for Genome Sciences (IGS) careers section. The header features the IGS logo and the text "Growth opportunities at a cutting edge Institute". The navigation menu includes links for Home, About, News & Events, Research, Resource Centers, People, Careers (which is currently selected), and Education. A sidebar on the left lists links for Introduction, Affiliate Faculty Membership, Faculty, Hiring Process, Internship Program, Post Doctoral Fellows, and Staff. The main content area displays a photograph of laboratory glassware and a computer monitor showing a bioinformatics interface. A sub-header "Careers – Introduction" is visible. Text on the page describes the institute's mission to advance careers in genomics and its focus on large-scale sequencing and analysis. It also mentions collaborations with other University of Maryland schools and global research centers. A video thumbnail in the bottom right corner shows people working in a lab, with a callout asking if viewers are interested in more information about a bioinformatics career at IGS.