

GSC Groups

Analysis and Annotation

Technology Development

Informatics

Assembly

Outreach

Mapping

Production

Prefinishing

Finishing

GSC Group Designations

Added Value Activities

Analysis and Annotation

Technology Development

Informatics

Assembly

Outreach

Pipeline Activities

Mapping

Production

Prefinishing

Finishing

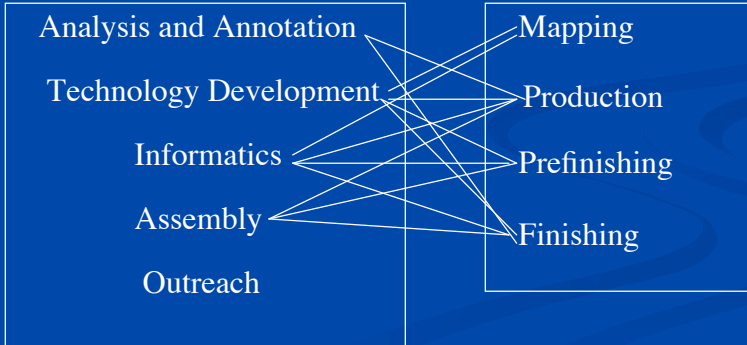
GSC Relationships

Added Value Activities

Analysis and Annotation
Technology Development
Informatics
Assembly
Outreach

Pipeline Activities

Mapping
Production
Prefinishing
Finishing



GSC Units of Measurement

Cost per Sequence Read

Mapping
Production

Cost per base pair

Analysis and Annotation
Prefinishing
Finishing

Others

Informatics Technology Development
Assembly Outreach

Reagents and salaries are charged directly to each group at the GSC when they are ordered or hired.

Cost per Sequence Read

Mapping
Production

Cost per base pair

Analysis and Annotation
Prefinishing
Finishing

Others

Informatics Technology Development
Assembly Outreach

The percent of effort from each support group is divided among the groups receiving their services.

Cost per Sequence Read

Mapping
Production

Cost per base pair

Analysis and Annotation
Prefinishing
Finishing

Others

Informatics Technology Development
Assembly Outreach

Total amount spent for each activity is divided by the unit of measurement, reads or bases.

Cost per Sequence Read

$\frac{(\text{Mapping} + 5\% \text{ Techd} + 10\% \text{ informatics})}{\text{Number of BAC end Reads}}$
 $\frac{(\text{Production} + 90\% \text{ TechD} + 40\% \text{ informatics})}{\text{Number of production reads}}$

Cost per base pair

Analysis and Annotation
 Prefinishing
 Finishing

Others

Informatics Technology Development
 Assembly Outreach

GSC Production Cost per Read

<u>2004</u>	<u>2005</u>	<u>Category</u>
45%	36%	Amortized Equipment
20%	21%	Indirect cost
10%	12%	Labor
2%	7%	Service Contracts
23%	24%	Supplies

GSC Production Cost per Read, Cost per Production Base Pair, and Amortized Equipment

Assume 1 sequencer and \$1,000 amortized equipment per machine:

13 runs/day @ 770 base pairs = 1,248 reads/day = 960,960 bases/day

\$0.80 AE cost/read or \$1.04 AE cost/kb

18 runs/day @ 680 base pairs = 1,728 reads/day = 1,175,040 bases/day

\$0.57 AE cost/read or \$0.85 AE cost/kb

24 runs/day @ 580 base pairs = 2,304 reads/day = 1,336,320 bases/day

\$0.43 AE cost/read or \$0.74 AE cost/kb

NHGRI Quarterly Progress Report for Production

Progress Report for Large-Scale Sequencing Grants: Individual Organism Report

Name of Center: Washington University in St. Louis

Use the upper tables to detail the sequencing strategies and activities funded for the three month period from November 1, 2004 to January 31, 2005. Use the lower tables to account for the costs of these activities. The shaded cells are pre-set to zero.

Organism: *Neisseria* (please specify)

IA. Sequence Production: Design Sequencing (Whole Genomes)	Cumulative total as of October 31, 2004			Month 1 (November 2004)			Month 2 (December 2004)			Month 3 (January 2005)			Three month total for November 1, 2004 - January 31, 2005			Cumulative total as of January 31, 2005		
	en. vs. planned	% vs. planned	other	en. vs. planned	% vs. planned	other	en. vs. planned	% vs. planned	other	en. vs. planned	% vs. planned	other	en. vs. planned	% vs. planned	other	en. vs. planned	% vs. planned	other
PT reads generated	7,716,342	0.0	0	2,124,244	0.0	0	30,424	30,244	0.0	4,724	1,787,244	0.0	1,344	100,744	2,124,244	0.0	1,344	100,744
PT reads sequenced	4,838,436	0.0	0	148,144	0.0	0	13,144	13,144	0.0	2,144	2,144	0.0	484	2,144	4,838,436	0.0	484	17,144
PT reads analyzed	48,144	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads generated	4,838,436	0.0	0	148,144	0.0	0	13,144	13,144	0.0	4,724	1,787,244	0.0	1,344	100,744	4,838,436	0.0	1,344	100,744
Cost of PT reads sequenced	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040	0.0	0	36,344	0.0	0	3,144	3,144	0.0	544	1,844	0.0	484	1,844	1,175,040	0.0	484	17,144
Cost of PT reads analyzed	4,838,436	0.0	0	48,144	0.0	0	48,144	48,144	0.0	48,144	48,144	0.0	48,144	48,144	48,144	0.0	48,144	48,144
Cost of PT reads with raw read data sent	1,175,040																	

NHGRI Quarterly Progress Report for Finishing

2. Products of Sequencing	Cumulative total as of October 31, 2004	Month 1 (November 2004)	Month 2 (December 2004)	Month 3 (January 2005)	Three month total for November 1, 2004 - January 31, 2005	Cumulative total as of January 31, 2005
A. Whole-Genome Activities						
PLEASE DESCRIBE THE PREVIOUS-ROUND ACTIVITIES AT YOUR CENTER IN A SEPARATE DOCUMENT						
Total # of pre-finished grade A reads	0	0	0	0	0	0
Total # of pre-finished grade A bp	0	0	0	0	0	0
Total # of pre-finished grade B reads	0	0	0	0	0	0
Total # of pre-finished grade B bp	0	0	0	0	0	0
Total # of whole genome finished top	0	0	0	0	0	0
B. BAC-based Activities						
Total # of pre-finished grade A reads	0	0	0	0	0	0
Total # of pre-finished grade A bp	0	0	0	0	0	0
Total # of pre-finished grade B reads	0	0	0	0	0	0
Total # of pre-finished grade B bp	0	0	0	0	0	0
Total # of BAC-based finished top	0	0	0	0	0	0
3. Data deposits:						
a. Traces						
Total # of depositions in the NCBI trace repository or made otherwise publicly available	6,945,983	0	12,853	2,673,670	299,223	9,632,329
b. Finished sequence						
Total phase_2 sequence bp deposited in GenBank	0	0	0	0	0	0
Total finished bp deposited in GenBank	0	0	0	0	0	0
Non-redundant finishedbp deposited in GenBank	0	0	0	0	0	0
4. Coverage:						
Cumulative total as of October 31, 2004	Month 1 (November 2004)	Month 2 (December 2004)	Month 3 (January 2005)	Three month total for November 1, 2004 - January 31, 2005	Cumulative total as of January 31, 2005	
Clone coverage12	16,952,109	1,316,466,697	6,181,729,697	95,799,9597	12,288,665	29,180,205
Sequence coverage	1,550,519,643	6,020,597,517	6,090,704,742	6,802,020,106	6,910,711,099	21,932,324,4

Production Information Captured in the NHGRI Progress Report

1A. Sequence Production: Shotgun Sequencing Whole Genome	Cumulative total as of October 31, 2004				Month 1 (November 2004)			
	sm. ins. plasmid1	lg. ins. plasmid1	fosmid	BAC end	sm. ins. plasmid1	lg. ins. plasmid1	fosmid	BAC end
# of reads attempted	7,756,592	0	0	215,424	0	0	0	32,448
# of successful reads2	6,895,806	0	0	99,126	0	0	0	19,747
Pass rate3	88.9%			46.0%				60.9%
Average read length4	669.5	0.0	0.0	532.2	0.0	0.0	0.0	544.7
Total # of Q20 bases5	4,616,742,117	0.0	0.0	52,754,857	0	0	0	10,755,511
# of clones with two successful end reads6	3,177,884	0	0	36,391	0	0	0	7,089
Total # of clones attempted7	4,020,880	0	0	107,701	0	0	0	16,224
Paired end rate13	92%	0%	0%	73%	0%	0%	0%	72%

Finishing Information Captured in the NHGRI Progress Report

3. Data deposits:	Cumulative total as of October 31, 2004	Month 1 (November 2004)
a. Traces		
Total # of depositions in the NCBI trace repository or made otherwise publicly available	6,945,983	0
b. Finished sequence		
Total phase_2 sequence bp deposited in GenBank	0	0
Total finished bp deposited in GenBank	0	0
Non-redundant finishedbp deposited in GenBank	0	0

Production and Finishing Statistics are used to calculate costs

PRODUCTION ONLY Costs should include: production-related informatics (including "routine" assembly, annotation and data submission); administrative; sequence production methods improvement aimed at increasing throughput, quality, and/or decreasing costs.

Sequence Production: Shotgun Sequencing	WHOLE GENOME				BAC-based			FOSSIL-based			OTHER	
	am. inv. planned*	% inv. planned*	actual	SAC and	am. inv. planned*	% inv. planned*	other (specify)	am. inv. planned*	% inv. planned*	other (specify)	estimate (specify)	estimate (specify)
Number of Q200 bases reported above	0.317,945,538	0	0.798,597	0.7948794.8	441,992,034	0	0	0	0	0	11,928,487	0
ACTIVITY COST TOTALS:	\$ 10,369,791	\$ -	\$ 28,826	\$ 111,890	\$ 620,817	\$ -	\$ -	\$ -	\$ -	\$ -	\$ 21,420	\$ -
Average Cost per shotgun read and Q20 reported for PRODUCTION ONLY activities	0.001238453		0.000219146	0.001098429	0.00149455						0.001798006	
Average Cost per read for FRODOING or Middlesex Post FINISHING activities (include the "Activity Cost Total" divided by total # reads or bases; this will be calculated automatically)												0.000879137
Library Making Costs Please enter funds used for producing libraries			523,269									
Three Month Total Costs (Production sequencing, pre-finishing, finishing and library construction) (summed from above)												13,421,128
Grand Total Three Month Costs (summed)												17,300,823
*Your "Three Month Total Cost" will be calculated according to the costs you present above, as well as the costs you report for technology development, bioinformatics, mapping, and other activities, which you fill in on the attached sheet. Please see:												
Cost Category	\$	%										
Contract Expenses (3 yrs)	1,884,588.73	10%										
Labor	3,305,789.07	19%										
Materials & Reagents	1,585,268.25	9%										
Sub-contracts	58,898.00	0%										
Travel	13,800.18	0%										
Other Direct Costs	535,417.45	3%										
Travel Cost Reimburse	13,980,688.91	81%										
Indirect Costs	3,307,179.84	19%										
Grand Total Three Month Costs	17,300,823	100%										
Explanation for reference	Total Grand Award amount reflects actual dollars. Grand Total Three Month Costs include amortized equipment calculations that are not actual expenses.											

The report also captures other activity costs. These costs are used to determine what percentage of the budget is used to support non-sequence related activities

Progress Report for Large-Scale Sequencing Grants: Other Activity Costs 0

Name of Center: _____

TOTAL COSTS: All funds used for all large-scale sequencing activities for the three-month period covered by this report (same as Grand Total Three Month Costs - Production Costs - all orgs spreadsheet).	\$17,300,829
---	---------------------

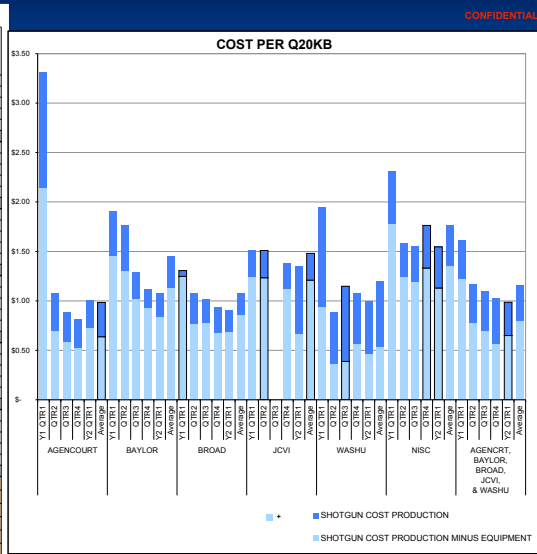
All costs listed below are those not attributed to sequence production (shotgun and finishing) described on the previous worksheet.

Category	Total Cost	% of Total Costs
TECHNOLOGY DEVELOPMENT COSTS (includes all costs spent on adaptation of new technologies developed elsewhere; implementation/testing of new protocols, etc.; excluded are costs spent on methods improvement aimed at increasing throughput, quality, and/or de	\$ 640,894	3.70%
BIOINFORMATICS DEVELOPMENT COSTS (includes all costs related to the development of improved bioinformatics methods, e.g. assembly algorithms, annotation software and development of production software; excluded are costs associated with "routine" assembly)	\$ 1,640,481	9.53%
MAPPING COSTS (includes all costs related to the generation of fingerprint maps)	\$ 803,739	4.65%
OTHER COSTS (include all other costs for which NHGRI funds are used which are not accounted for in the activities listed on this sheet or elsewhere (e.g. community coordination costs and outreach costs); this does not include funds for the Minority Action)	\$ 775,583	4.48%

DESCRIPTION OF ACTIVITIES: Please complete the included Word document describing all the activities associated with the four categories on this worksheet whose funding you have accounted for above. Please describe the activities and their corresponding co

The information provided to NHGRI is used to compare sequencing centers and monitor costs.

LARGE-SCALE SEQUENCING CENTERS PER Q20 KLOBASE COSTS				
CENTER	PERIOD	SHOTGUN COST PRODUCTION MINUS EQUIPMENT	SHOTGUN COST PRODUCTION	COST PERCENTAGE DEDUCTED TO EQUIPMENT
AGENCOURT	Y1 QTR1	\$ 2.14	\$ 3.31	35%
	QTR2	\$ 0.70	\$ 1.08	35%
	QTR3	\$ 0.59	\$ 0.89	24%
	QTR4	\$ 0.52	\$ 0.81	36%
	Y2 QTR1	\$ 0.73	\$ 1.00	28%
Average	\$ 0.66	\$ 0.86	29%	
BAYLOR	Y1 QTR1	\$ 1.45	\$ 1.90	23%
	QTR2	\$ 1.31	\$ 1.77	26%
	QTR3	\$ 1.02	\$ 1.29	21%
	QTR4	\$ 0.93	\$ 1.11	16%
	Y2 QTR1	\$ 0.84	\$ 1.08	22%
Average	\$ 1.14	\$ 1.46	22%	
BROAD	Y1 QTR1	\$ 1.25	\$ 1.31	4%
	QTR2	\$ 0.77	\$ 1.08	29%
	QTR3	\$ 0.78	\$ 1.02	23%
	QTR4	\$ 0.68	\$ 0.84	28%
	Y2 QTR1	\$ 0.69	\$ 0.90	24%
Average	\$ 0.86	\$ 1.08	21%	
JCVI	Y1 QTR1	\$ 1.24	\$ 1.61	18%
	QTR2	\$ 1.23	\$ 1.51	19%
	QTR3	\$ 1.13	\$ 1.39	19%
	QTR4	\$ 0.86	\$ 1.35	51%
	Average	\$ 1.21	\$ 1.48	18%
WASHU	Y1 QTR1	\$ 0.94	\$ 1.94	52%
	QTR2	\$ 0.36	\$ 0.89	29%
	QTR3	\$ 0.39	\$ 1.15	66%
	QTR4	\$ 0.56	\$ 1.07	48%
	Y2 QTR1	\$ 0.46	\$ 1.00	54%
Average	\$ 0.54	\$ 1.03	50%	
NISC	Y1 QTR1	\$ 1.78	\$ 2.31	23%
	QTR2	\$ 1.25	\$ 1.59	21%
	QTR3	\$ 1.20	\$ 1.56	23%
	QTR4	\$ 1.33	\$ 1.76	24%
	Y2 QTR1	\$ 1.13	\$ 1.55	27%
Average	\$ 1.38	\$ 1.77	28%	
AGENCRT, BAYLOR, BROAD, JCVI, & WASHU	Y1 QTR1	\$ 1.22	\$ 1.81	25%
	QTR2	\$ 0.78	\$ 1.17	33%
	QTR3	\$ 0.70	\$ 1.09	36%
	QTR4	\$ 0.57	\$ 1.03	45%
	Y2 QTR1	\$ 0.65	\$ 0.89	34%
Average	\$ 0.79	\$ 1.13	31%	



NHGRI costs are used to figure the costs for other collaborators who access the GSC for sequencing work - like Bio 4342

The GSC provides: Production Sequencing, Finishing, and Analysis to collaborators at the university and outside the university

Production Sequencing Cost by Read Number in 1 direction					
Total Number of 36 well trays	1	1	1	1	1
Total Number of 96 well trays	1	1	1	1	1
Total Number of Subclones	1	1	1	1	1
One End	1	1	1	1	1
Total Number of Reads	1	1	1	1	1
Total Number of Peeking Reads	1	1	1	1	1
Percent Success	90%	90%	90%	90%	90%
Cost per read	\$ 0.70	\$ 0.70	\$ 0.70	\$ 0.70	\$ 0.70
Total Cost for Sequencing Reads	\$ 0.70	\$ 0.70	\$ 0.70	\$ 0.70	\$ 0.70
Total Cost for Sequencing Reads	\$ 0.70	\$ 0.70	\$ 0.70	\$ 0.70	\$ 0.70
Production Sequencing Cost by Sequencing Coverage					
Production - Development Costs					
Development - Paired Libraries	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00
Development - Cost/Peaked Library	\$ -	\$ -	\$ -	\$ -	\$ -
Development - Total Cost for Paired Libraries	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00
Development - WGS Libraries	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00
Development - Cost/WGS Library	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00
Development - Total Cost for WGS Libraries	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00
Development - Clone-based Libraries (MAC/Barcode)	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00
Development - Cost/Clone-based Library	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00
Development - Total Cost for Clone-based Libraries	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00
Development - Total Cost/Read	\$ 2.00	\$ 2.00	\$ 2.00	\$ 2.00	\$ 2.00
Development - Total Cost/Read	\$ 2.00	\$ 2.00	\$ 2.00	\$ 2.00	\$ 2.00
Development - Total Development Costs	\$ 2,000.00	\$ 2,000.00	\$ 2,000.00	\$ 2,000.00	\$ 2,000.00
Library Construction Costs					
Production - Paired Libraries	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00
Production - Cost/Peaked Library	\$ -	\$ -	\$ -	\$ -	\$ -
Total Production Paired Library Cost	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00	\$ 1,000.00
Production - WGS Libraries	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00
Production - Cost/WGS Library	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00
Total Production WGS Library Cost	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00	\$ 600.00
Production - Clone-based Libraries	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00
Production - Cost/Clone-based Library	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00
Total Production Clone-based Library Cost	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00	\$ 300.00
Total Production Library Construction Cost	\$ 2,000.00	\$ 2,000.00	\$ 2,000.00	\$ 2,000.00	\$ 2,000.00
Shotgun coverage - 4x for 7 kb					
Genome Size or Size of Clone	7 kb	7 kb	7 kb	7 kb	7 kb
Insert Size	7 kb	7 kb	7 kb	7 kb	7 kb
Avg. read length (bp)	3,500	3,500	3,500	3,500	3,500
Desired Seq coverage (X)	4	4	4	4	4
Resulting physical coverage	14	14	14	14	14
Number of reads in CG	14	14	14	14	14
Number of reads attempt	14	14	14	14	14
Cost per read	\$ 0.86	\$ 0.86	\$ 0.86	\$ 0.86	\$ 0.86
Project Production cost	\$ 12.04	\$ 12.04	\$ 12.04	\$ 12.04	\$ 12.04
Shotgun coverage - 2x for 10 kb					
Genome Size or Size of Clone	10 kb	10 kb	10 kb	10 kb	10 kb
Insert Size	10 kb	10 kb	10 kb	10 kb	10 kb
Avg. read length (bp)	3,500	3,500	3,500	3,500	3,500
Desired Seq coverage (X)	2	2	2	2	2
Resulting physical coverage	7	7	7	7	7
Number of reads in CG	7	7	7	7	7
Number of reads attempt	7	7	7	7	7
Cost per read	\$ 0.86	\$ 0.86	\$ 0.86	\$ 0.86	\$ 0.86
Project Production cost	\$ 6.02	\$ 6.02	\$ 6.02	\$ 6.02	\$ 6.02
Paired shotgun coverage					
Genome Size or Size of Clone	7 kb	7 kb	7 kb	7 kb	7 kb
Insert Size	7 kb	7 kb	7 kb	7 kb	7 kb
Avg. read length (bp)	3,500	3,500	3,500	3,500	3,500
Desired Seq coverage (X)	2	2	2	2	2
Resulting physical coverage	4	4	4	4	4
Number of reads in CG	4	4	4	4	4
Number of reads attempt	4	4	4	4	4
Cost per read	\$ 0.86	\$ 0.86	\$ 0.86	\$ 0.86	\$ 0.86
Project Production cost	\$ 3.44	\$ 3.44	\$ 3.44	\$ 3.44	\$ 3.44
Project Totals					
Each project total	\$ 12.04	\$ 12.04	\$ 12.04	\$ 12.04	\$ 12.04

At any one time the GSC has several ongoing projects for both NHGRI and outside collaborators

All of the projects are tracked weekly to ensure they move quickly through the pipeline

This is the current list of projects.

The GSC Production Schedule for NHGRI

Project	Total	Target	Start	End	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	
Project 1	100	100	2010	2011	100																					
Project 2	100	100	2010	2011	100																					
Project 3	100	100	2010	2011	100																					
Project 4	100	100	2010	2011	100																					
Project 5	100	100	2010	2011	100																					
Project 6	100	100	2010	2011	100																					
Project 7	100	100	2010	2011	100																					
Project 8	100	100	2010	2011	100																					
Project 9	100	100	2010	2011	100																					
Project 10	100	100	2010	2011	100																					
Project 11	100	100	2010	2011	100																					
Project 12	100	100	2010	2011	100																					
Project 13	100	100	2010	2011	100																					
Project 14	100	100	2010	2011	100																					
Project 15	100	100	2010	2011	100																					
Project 16	100	100	2010	2011	100																					
Project 17	100	100	2010	2011	100																					
Project 18	100	100	2010	2011	100																					
Project 19	100	100	2010	2011	100																					
Project 20	100	100	2010	2011	100																					
Project 21	100	100	2010	2011	100																					
Project 22	100	100	2010	2011	100																					
Project 23	100	100	2010	2011	100																					
Project 24	100	100	2010	2011	100																					
Project 25	100	100	2010	2011	100																					
Project 26	100	100	2010	2011	100																					
Project 27	100	100	2010	2011	100																					
Project 28	100	100	2010	2011	100																					
Project 29	100	100	2010	2011	100																					
Project 30	100	100	2010	2011	100																					

The schedule of reads are turned into a report with Q20 bases for NHGRI. Goals are evaluated quarterly.

NHGRI Large-Scale Sequencing Centers: PROJECTED Q20 base pairs PER MONTH (in MILLIONS)

Organism	Wilson Strategy	Overall Project	Center Seq. Cov.	PROJECTED Q20 base pairs PER MONTH (in MILLIONS)																			
				NOV	DEC	JAN (2004)	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEPT	OCT	NOV	DEC	JAN (05)	FEB	MAR	APR	MAY	JUN
<i>Quarterm</i>	BAC/WGS	EX	EX	21	21	25	33	118	85	99	3	27	1	1	1	1	8	9	28	14	14	14	
<i>Homo sapiens</i>	BAC/WGS	EX	EX	89	101	75	105	44	103	119	84	33	17	88	81	49	47	40	171	179	170	179	
<i>Gallus gallus gallus</i>	BAC/WGS	EX	EX	163	179	17	6	1									7	10	38	37			
<i>Pan troglodytes</i>	BAC/WGS	EX	EX	229	118	807	974	407	2158	1548	828	6	45	105	33	105	85	30	35	34	48	48	
<i>Drosophila yakuba</i>	WGS	EX/1X	EX/1X	691	695	192	36																
<i>Drosophila simulans</i>	WGS	EX	EX	121	2	189	177	166	127	388	19	173	17	283									
<i>Xenopus</i>	BAC ends	0.2X	0.2X																				
<i>Macaca mulatta</i>	WGS	EX	2X						7	114	2241	2440	220	17	20	11	18	1928	879				
<i>Schmidtea mediterranea</i>	WGS	EX	EX	798	48		87	30	1	10	1	7	2228	1984	1764	6			408				
<i>Ornithorhynchus anatinus</i>	WGS/BAC	EX	EX			3			68	48	6	20	1	877	589	2667	2093	1177	1915	2393	3087	198	
<i>Caenorhabditis remanei</i>	WGS	EX	EX					972	8						27	1			1	18	7		
<i>Caenorhabditis elegans</i>	WGS	EX	EX																				
<i>Caenorhabditis s. sp. 09/2801</i>	WGS	EX	EX																				
<i>Pristionchus pacificus</i>	WGS	EX	EX																	1647	122	7	
<i>Lamprocy</i>	WGS/BAC	EX	EX						24											2	259	261	261
<i>Simulium oragutan</i>	IS	EX	3x																				
<i>Oryzichia latallax</i>	IS	EX	EX																				
<i>Saccharomyces</i>	WGS	EX	EX																				
<i>Biomphalaria alabrate</i>	IS																						
<i>Physoarum polycephalum</i>	IS																						
<i>Monosiga ovata</i>	IS																						
<i>Callithrix jacchus</i>	IS																						
<i>Microbodus murinus</i>	IS																						
<i>Hydrobatas concolor</i>	IS																						
<i>Acropora millepora (coral)</i>	IS																						
<i>Periliss lobata</i>	IS																						
<i>Acropora palmata</i>	IS																						
<i>Symbiodinium</i>	IS																						
<i>Convolvula palustris</i>	IS																						
<i>Lizard Comparison</i>	IS																						
<i>Unnamed Organism/Testing</i>	IS																						

GSC Production

<u>2004</u>	<u>2005</u>	
3.1M	7.0M	Reads Per Month
770 bp	680 bp	Read Length
92%	92%	Percent Success
4	4	Managers
10	9	Coordinators
40	50	Technicians

Production Goals

New Sequencing Technologies

Decrease labor and supply costs

Lower amortized equipment costs

Production Projects

Alignment with prefinishing

Mutational Profiling

Increase in special projects

Refine database tools

Loss Prevention/Monitoring

GSC Prefinishing

40 MB per month

400 clones per month

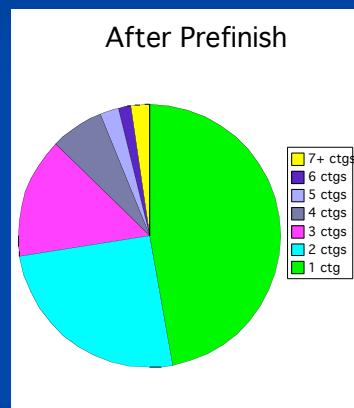
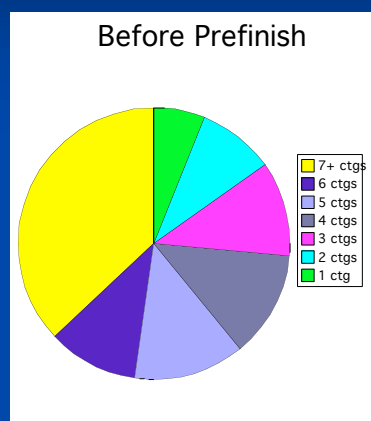
Prefinished 12 genome projects

1 manager

1 coordinator

8 technicians

Prefinishing Data



Prefinishing Goals

Increase Automation
Complete Oracle Integration
Increased role of freezer core
Tools for whole genome projects
Alignment of sorters with
pathfinders

GSC Finishing

50 MB per month
270 clones per month
Finished 6 whole genome projects
4 managers
1 assistant coordinators
40 finishers

Finishing Goals

Integration into Oracle db
Implementation of automation
Organization of finishing groups
Improved draft sequence
Map and sequence integration
Decrease cost of a finished base