



SCGC Service Description

For further information, go to scgc.bigelow.org

Updated June 1, 2018

Single cell genomics unveils the genomic blueprints of the most fundamental units of life without the need for cultivation. This is a powerful approach to analyze biochemical properties, evolutionary histories and biotechnological potential of uncultured microorganisms, which constitute over 99% of biological diversity on our planet. Single cell genomics is also emerging as a revolutionary technology in the studies of cancer, autoimmune diseases and hereditary disorders. SCGC offers a comprehensive suite of single cell and single viral particle genomics services, from cell/viral separation to genome sequencing and bioinformatics.

Single cell genomics consists of a series of integrated processes, starting with appropriate collection and preservation of environmental samples, followed by physical separation, lysis, and whole genome amplification of individual cells, then proceeding to either targeted loci or whole genome sequencing and sequence interpretation (1). SCGC offers a comprehensive suite of single cell genomics services, from single cell separation through genome sequencing and bioinformatics. SCGC also provides advice on environmental sample collection and storage protocols ([please visit Preparation and Shipment for detailed shipping instructions](#)) and post-sequencing analyses (customized services). The [general methods that are currently employed by SCGC](#) have been described previously.

[To view examples of SCGC service results, click here.](#) A detailed description of SCGC services is available for download as a PDF file.

Table 1. List of SCGC services. For a detailed description, see text below.

Service	Cat. #	Unit	Unit price*
SINGLE CELL ISOLATION AND WHOLE GENOME AMPLIFICATION			
SAG Generation 2	S-201	384-well plate	\$3,800
SINGLE CELL DNA SEQUENCING AND BIOINFORMATICS			
LoCoS – Low Coverage Sequencing	S-202	384-well plate	\$15,000
Post-LoCoS WGS of prokaryotes**	S-203	Up to 100	\$15,000
Prokaryote SAG WGS, small batch	S-015	10 SAGs	\$9,900
Prokaryote SAG WGS, medium batch	S-016	48 SAGs	\$29,900
Eukaryote SAG WGS	S-018	flexible	Req. a
ADDITIONAL SERVICES			
SAG re-arraying	S-105	96-well plate	\$450
Sample cryoprotectant glyTE	S-019	10 mL	\$100
Consultation	S-011	1 hour	\$260
Customized Services	S-100		Req. a

*All prices are in US dollars (USD). Service fees do not include shipping costs. 15% discount is applied if more than 10 same items (catalog number) are ordered simultaneously. 15% discount is also applied if any combination of SCGC

services that exceed \$200,000 in value are ordered simultaneously. 20% discount is applied on all services for Bigelow Laboratory in-house projects.

** To be eligible for this service, all Illumina libraries must originate from a single SAG plate on which LoCoS (S-202) has already been performed. Furthermore, SAGs for which insufficient Illumina library quantities are produced during LoCoS (<0.2% of reads in a LoCoS set) are not eligible for this service.

NOTE: As of January 1, 2018, we no longer accept new requests for legacy services “SAG Generation 1” and PCR-based screens (S-101, S-102, S-103, S-104 and S-106). For a limited time, these services may be arranged for already ongoing SCGC projects.

COST ESTIMATE: CASE STUDY

Let’s assume that an external SCGC customer wants to a) generate one 384-well plate of prokaryote single amplified genomes (SAGs); b) perform a low coverage sequencing on this plate; and 3) perform deeper sequencing on 100 SAGs from the same plate. Fees for this work are summarized in Table 2. Deliverables include:

1. One 384-well microplate containing SAGs with associated flow cytometry and gDNA amplification kinetics data
2. Annotated genome sequences and QC of 100 deep-sequenced SAGs and associated metadata.
3. Annotated genome sequences and QC of the remaining 217, LoCoS-sequenced SAGs and associated metadata.

Table 2. Cost estimate for a hypothetical SCGC project.

Service	Cat. #	Unit	Price per unit	# of units	Amount
SAG Generation 2	S-201	384-well plate	\$3,800	1	\$3,800
LoCoS – Low Coverage Sequencing	S-202	384-well plate	\$15,000	1	\$15,000
Post-LoCoS WGS of prokaryotes	S-203	Up to 100 SAGs	\$15,000	1	\$15,000
Total					\$33,800

DETAILED SERVICE DESCRIPTION

SAG generation 2 (S-201)

Single amplified genomes (SAGs) are whole genome amplification products obtained from individual cells or individual particles of DNA viruses. They can be used in similar ways to DNA extracts from pure cultures. The S-201 service includes separation of individual cells or DNA-containing viral particles into wells of a 384-well plate by fluorescence-activated cell sorting (FACS), followed by cell lysis with KOH and genomic DNA amplification by WGA-X, a new method developed at SCGC (2). Compared to the earlier, multiple displacement amplification technique (3), WGA-X improves average genome recovery from individual cells and viral particles, with most notable enhancements observed in SAGs with high G+C content. Deliverables of S-201 include: a) one 384-well microplate containing single cell or viral particle WGA-X products, 10 µL per well, usually averaging 1 microgram gDNA per well; b) index FACS data files; and c) WGA-X kinetics data obtained by measuring fluorescence of a DNA stain. One row and three columns of positive and negative controls are included in each 384-well microplate (Figure 1). For a standard S-201 service fee to be applicable, the following conditions must be met:

1. Cells or viruses are in aquatic solution, are less than 100 micrometers in diameter, and are cryopreserved and shipped following SCGC recommendations ([see Preparation and Shipment on SCGC website](#)). Only particles <40 micrometers in diameter can be sorted in index FACS mode.

- Cell or viral sorting is based on SYTO-9 fluorescence (provided by SCGC), particle autofluorescence, or probes that are applied by SCGC customer prior to shipping to SCGC.
- Only one sample is used per microplate, following SCGC's standard plate setup (Figure 1).

Samples that have not been brought into aquatic suspension by SCGC customers may also be analyzed as a customized service. Please contact SCGC manager [Brian Thompson](#) for additional advice on how to produce SCGC-compatible samples in your own lab or for a feasibility assessment and a quote.

SAG generation success varies among samples and depends on many factors, such as: a) preservation of intact cells and gDNA prior to cell sorting; b) successful discrimination of cells and viral particles from other particles during FACS; and c) successful single cell lysis and DNA amplification (Figure 1). Please note that viral particles may be particularly difficult to discriminate from other DNA-containing particles in sub-cellular size range when working with non-aquatic samples.

[To view examples of SCGC service results, click here.](#)

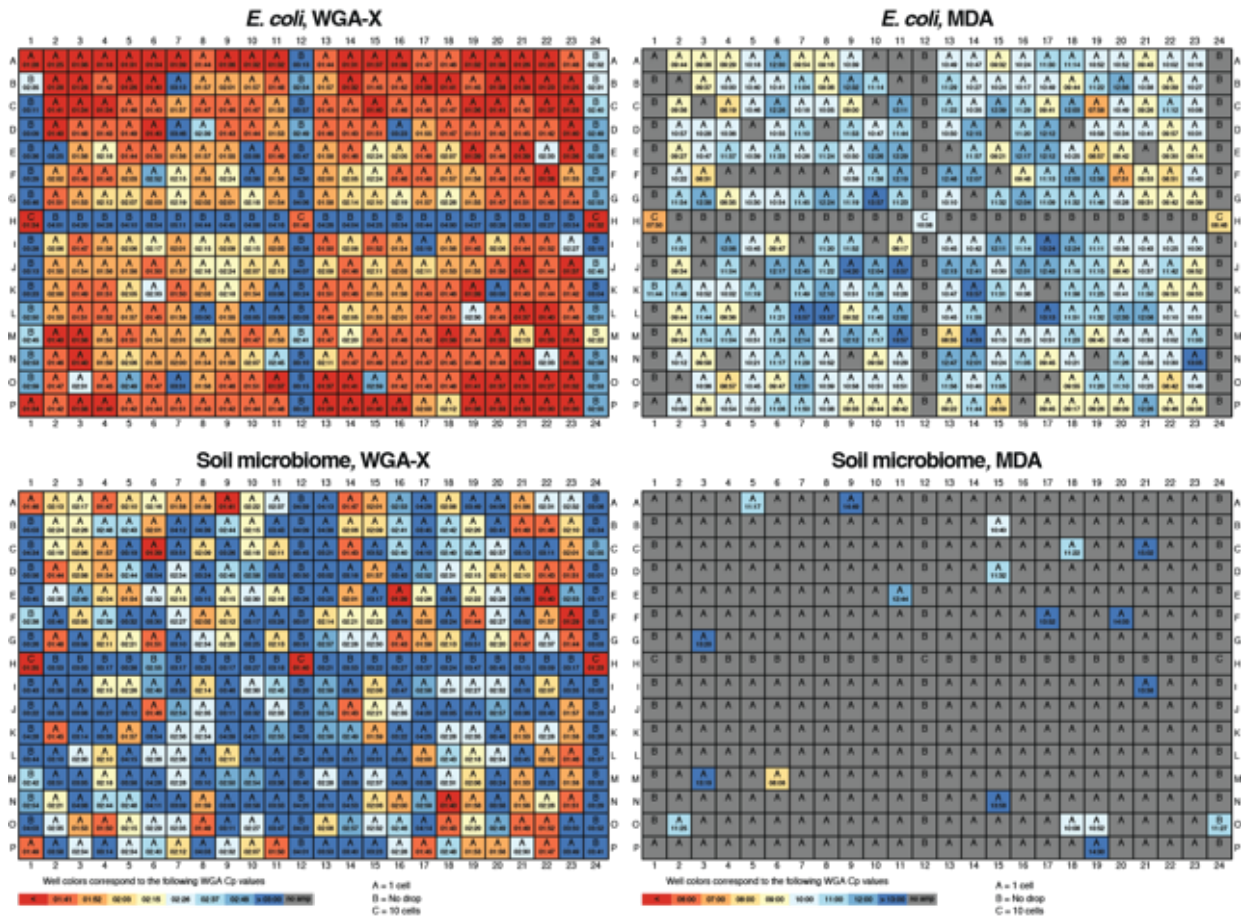


Figure 1. Examples of WGA-X and MDA kinetics in microplates with divergent DNA amplification success rates. The reaction Cp value indicates time (hours) required to reach half of the maximal DNA-SYTO-9 fluorescence. The Cp is inversely correlated with amplification speed and the initial quantity of gDNA that is available to polymerase.

LoCoS – Low Coverage Sequencing (S-202)

SCGC developed LoCoS in order to retrieve some genomic information from most SAGs on a 384-well plate at a minimal cost, currently under \$50 per genome. Deliverables include: a) original, 2×150 bp sequence reads (number per SAG is variable, average around 300,000), assembled contigs, assembly statistics, functional annotation by Prokka (14), taxonomic assignments based on checkM (15) and SSU rRNA gene phylogeny, and QC. Please note that the obtained sequencing depth is highly variable among SAGs and is not sufficient for high-quality genome recovery from most SAGs. Therefore S-202 should be considered primarily as an initial screen of SAGs, to be followed by deeper sequencing (e.g. SCGC services S-203, S-015, S-016 or S-018). Further information on LoCoS and research examples are available in Stepanauskas et al. (2017).

[To view examples of SCGC service results, click here.](#)

Post-LoCoS WGS of prokaryotes (S-203)

This service includes deeper sequencing, *de novo* assembly and functional annotation of up to 100 Illumina libraries that have already been produced, in sufficient quantity, by an S-202 service from a single SAG plate. The total number of 2×150 bp reads produced by this service is around 400 million and is distributed about equally among the selected SAGs. Thus, the sequencing depth of each SAG depends on the number of SAGs selected for this service. Larger number of reads per SAG usually results in better genome recovery, but this relationship is not linear (Figure 2).

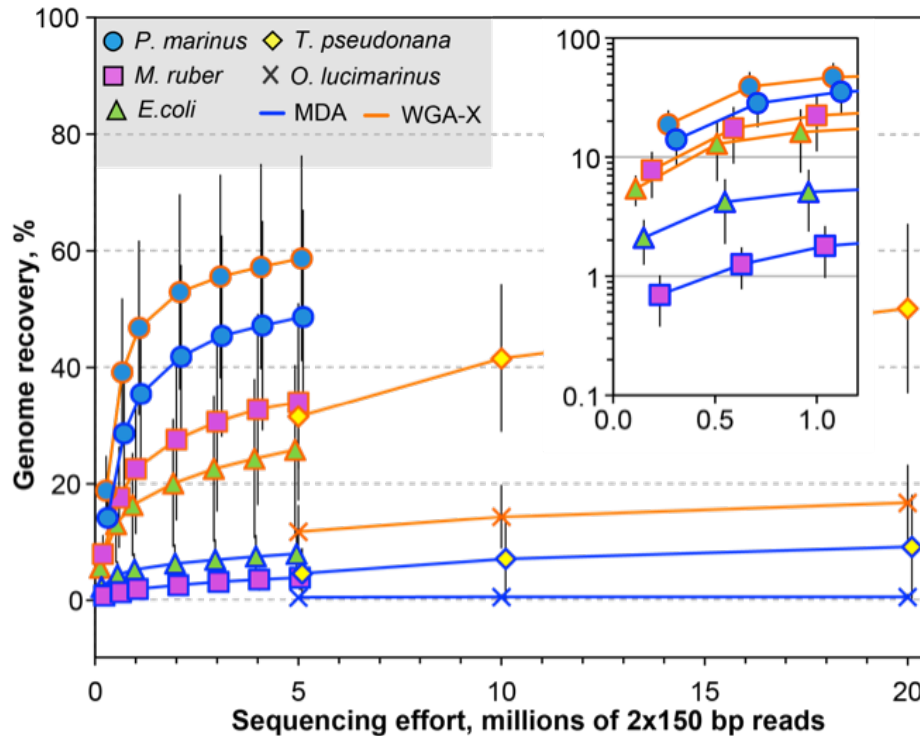


Figure 2. Genome recovery from SAGs of benchmark microorganisms as a function of sequencing effort. Paired-end reads of each bacterial SAG were randomly down-sampled to 0.1, 0.5, 1, 2, 3, 4, and 5 millions. Paired-end reads of each eukaryote SAG were randomly down-sampled to 5, 10 and 20 millions. Note that data markers are slightly offset on the horizontal scale, in order to better discriminate standard deviation bars among the treatments. The inset provides a better view of results at low read numbers. Displayed are averages and standard deviations, which were estimated from eight randomly selected SAGs in the case of bacterial SAGs, and from three SAGs with the lowest reaction C_p values in the case of eukaryote SAGs. From (2).

Please note that, in order to be eligible for this service, all Illumina libraries must originate from a single SAG plate on which LoCoS (S-202) has already been performed. Furthermore, SAGs for which insufficient Illumina library quantities are produced during LoCoS (<0.2% of all LoCoS reads from a SAG plate) are not eligible for this service. Other SAGs can be sequenced using SCGC services S-015, S-016 or S-018. Per customer's request, SCGC can generate lists of SAGs that meet S-203 eligibility criteria. Please note that, due to potential liquid handling inconsistencies, some SAGs may receive insufficient numbers of sequence reads. Due to its high-throughput, low-cost nature, service S-203 does not include free re-sequencing of such SAGs, unless the number of affected SAGs exceeds 10 per plate.

Please note that genome recovery from SAGs is variable and ranges from 0% to 100% (2). Incomplete genome recovery may be caused by diverse factors, such as DNA degradation due to viral infections or handling prior to sorting, incomplete cell lysis, DNA interactions with other cellular components, amplification and sequencing biases, and others. If multiple, closely related SAGs of interest are available, the following strategies may improve the probability of high genome recovery: a) sequencing SAGs with lowest WGA-X or MDA Cp values; and b) sequencing multiple, related SAGs for a population-level genome content analysis.

Prokaryote SAG WGS (S-015 and S-016)

These services include preparation of Illumina libraries, sequencing of at least 3 million 2×150 bp reads per SAG, *de novo* assembly and functional annotation and QC of each SAG. Deliverables include original sequence reads, assembled contigs, assembly statistics, annotation by Prokka (14), taxonomic assignments based on checkM (15) and SSU rRNA gene phylogeny, and the output of various QC steps. In difference to S-203, each service request may contain SAGs from multiple plates and there is no requirement for S-202 being done first.

Please note that genome recovery from SAGs is variable and ranges from 0% to 100% (2). Incomplete genome recovery may be caused by diverse factors, such as DNA degradation due to viral infections or handling prior to sorting, incomplete cell lysis, DNA interactions with other cellular components, amplification and sequencing biases, and others. If multiple, closely related SAGs of interest are available, the following strategies may significantly improve the probability of high genome recovery: a) sequencing SAGs with lowest WGA-X or MDA Cp values; and b) sequencing multiple, related SAGs for a population-level genome content analysis.

Eukaryote SAG WGS (S-018)

This service includes preparation of Illumina libraries, sequencing and *de novo* assembly of eukaryote SAGs. Deliverables include original sequence reads, assembled contigs, assembly statistics, taxonomic assignments based on SSU rRNA gene phylogeny, and QC. Genome annotation is not included in this service. If no greater sequencing depth than for prokaryotes is desired, then the fees are identical as for prokaryote SAG WGS (see S-202, S-203, S-015 and S-016). However, due to the highly variable and generally larger than prokaryote genome sizes of eukaryotes, greater sequencing effort may be needed for a comparable genome fraction to be recovered (Figure 2), in which case the fee for this service depends on the desired sequencing effort. [Please contact SCGC for a quote.](#)

As with prokaryotes (see above), the fraction of a genome recovered from each eukaryote SAG is variable (Figure 2). Benchmarking of SCGC's workflow on two marine algae demonstrated high data quality, with the frequency of assembly artifacts that is comparable to those in prokaryote SAGs (2). However, please note that SPAdes (16), which we use in *de novo* assemblies, has not been optimized for diploid organisms. Furthermore, we have not tested the performance of our *de novo* assembly pipeline on genomes larger than 32 Mbp and therefore cannot guarantee the provision of such assemblies.

SAG re-arranging (S-105)

SAGs are transferred from the original 384-well plates to 96-well plates. The SCGC customer defines transfer volumes, source wells, and destination wells. Prior to the transfer, the destination wells are pre-filled with 5-150 µL of either deionized water or 1x TE buffer, as specified by the SCGC customer. Deliverables include re-arranged SAGs in a 96-well plate.

Sample cryoprotectant glyTE (S-019)

For SCGC customers' convenience, we offer the preparation and shipment of SCGC's recommended sample cryoprotectant glyTE. Deliverables of this service include preparation and shipment of 10 mL of 10x concentrated glyTE. Please note that SCGC's recommended procedures for sample cryopreservation, including the recipe for the cryoprotectant glyTE, are posted on the SCGC website under [Preparation and Shipment](#).

Consultation (S-011)

Basic support is included in the pricing of all SCGC services. When SCGC customers require more extensive help with study design and/or data interpretation (greater than two hours per project), we may request compensation for the associated labor costs via collaborative research grants or consultant fees. We may also request co-authorship on resulting publications for those SCGC scientists who are providing substantial, project-specific intellectual input.

Customized services (S-100)

SCGC offers a wide range of customized services, e.g., non-standard cell sorting, non-SAG sequencing, bioinformatics support, method development, etc. Co-authorship on resulting publications may be requested for those SCGC scientists who are providing substantial, project-specific intellectual input. For more information, please contact SCGC manager Brian Thompson (bthompson@bigelow.org).

Note on SAG sequencing outside SCGC

Specialized laboratory and computational procedures are necessary in order to obtain high quality genomic sequences from SAGs. One challenge to consider is the cross-talk of multiplexed Illumina libraries (also known as "sample bleeding" and "index switching"):

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0120520>
<http://biorxiv.org/content/early/2017/04/09/125724>
<https://www.wired.com/2017/04/geneticists-fear-illumina-sequencers-may-distort-results/>
<https://sequencing.qcfail.com/articles/mixing-sample-types-in-a-flowcell-lane-generates-cross-contamination-artefacts/>
<http://enseqlopedia.com/2016/12/index-mis-assignment-between-samples-on-hiseq-4000-and-x-ten/>

De novo assemblies of single amplified genomes (SAGs) are especially sensitive to this issue. This is due to the fact that single cell whole genome amplification is highly uneven across the genome, and deep sequencing is typically employed to facilitate the recovery of the under-amplified genome regions (Figure 2). As a result, even a relatively small overall fraction of miss-assigned reads may form sizeable contigs that represent over-amplified regions of co-sequenced SAGs. In 2013, SCGC identified library cross-talk in MiSeq and HiSeq 2000 as a major problem and addressed it by computational decontamination of SAG assemblies. In summer 2014, we established in-house infrastructure and procedures for the preparation of dual-barcoded libraries and their sequencing with NextSeq 500, followed by optimized data processing. These measures have eliminated library cross-talk and the need for computational decontamination (2). To verify the efficacy of these solutions, SCGC benchmarks its entire workflow using SAGs of previously sequenced strains:

http://data.bigelow.org/~scgc/WGS_benchmark_data/

We encourage SCGC customers to take similar precautions and workflow validation when sequencing SAGs outside SCGC, in order to maintain data integrity. Also, please note that SCGC does not accept back any SAG plates that have been handled elsewhere, in order to prevent the risk of DNA contamination at our facility.

[To view examples of SCGC service results, click here.](#)