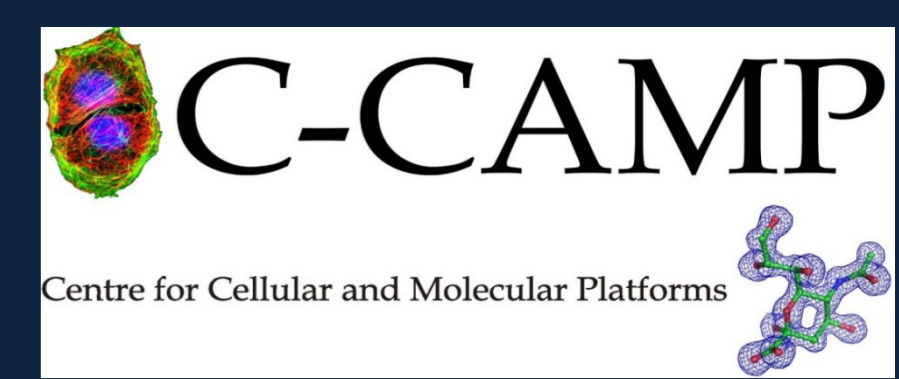


# Next-Generation Genomics Facility (Accelerating Life-Science Research)



**Malali Gowda\*, Chandana S, Heikham Russiachand, Pradeep H, Shilpa S, Ashwini M, Sahana S, Goutham Atla, Smita Jain, Nandini Arunkumar and Papri Banerjee.**

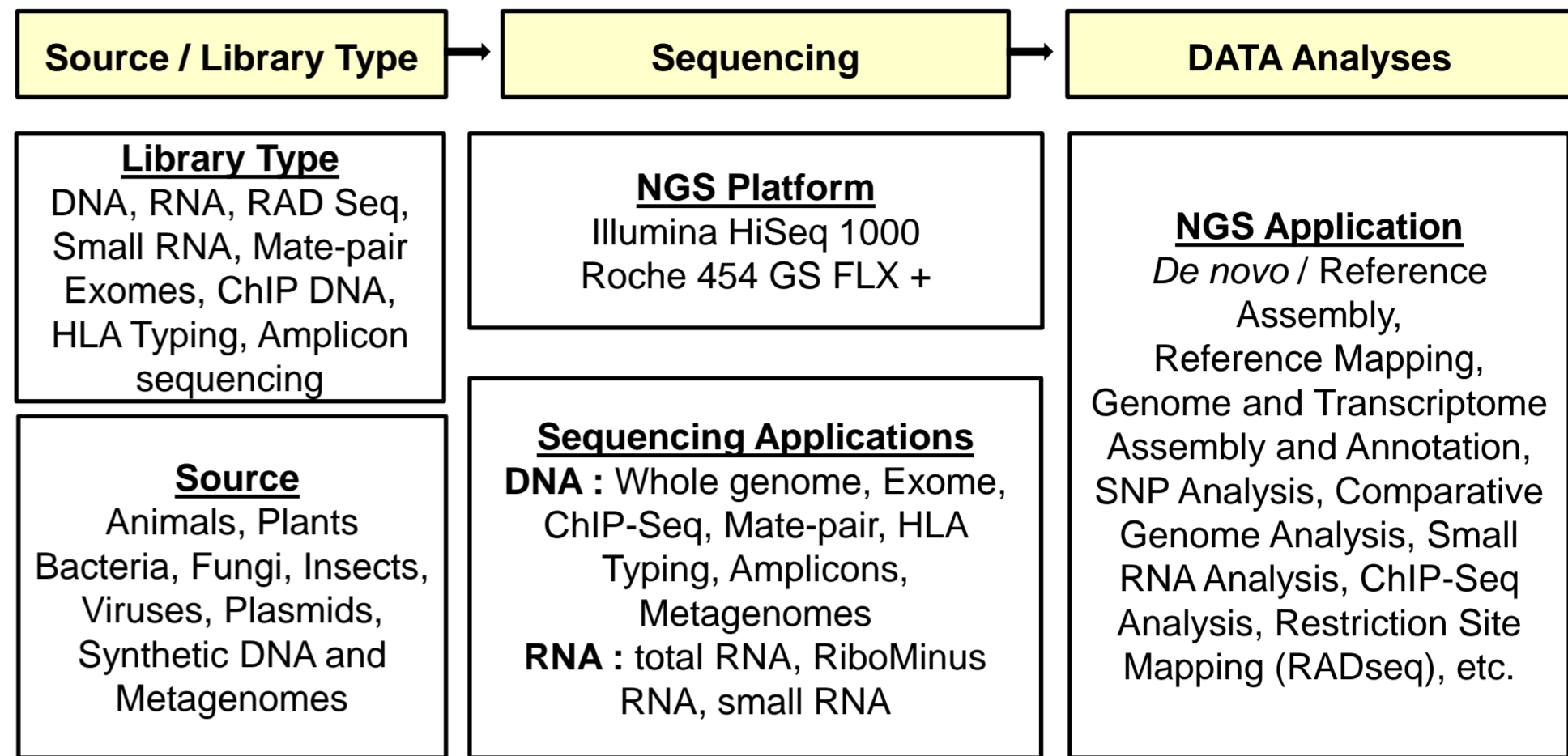
Centre for Cellular and Molecular Platforms (C-CAMP), National Centre for Biological Sciences, GKVK Post, Bangalore – 560 065

\*Email: [services@ccamp.res.in](mailto:services@ccamp.res.in), [malalig@ncbs.res.in](mailto:malalig@ncbs.res.in); Web: [www.ccamp.res.in/NGS-Genomics](http://www.ccamp.res.in/NGS-Genomics), Phone: 080-67185055

## Abstract

Next-Generation Sequencing (NGS; <http://www.genome.gov/12513162>) is a recent life-sciences technological revolution that allows scientists to decode genomes or transcriptomes at a much faster rate with a lower cost. Genomic-based studies are in a relatively slow pace in India due to the non-availability of genomics experts, trained personnel and dedicated service providers. Using NGS there is a lot of potential to study India's national diversity (of all kinds). We at the Centre for Cellular and Molecular Platforms (C-CAMP) have launched the Next Generation Genomics Facility (NGGF) to provide genomics service to scientists, to train researchers and also work on national and international genomic projects. We have HiSeq1000 from Illumina and GS-FLX Plus from Roche454. The long reads from GS FLX Plus, and high sequence depth from HiSeq1000, are the best and ideal hybrid approaches for de novo and re-sequencing of genomes and transcriptomes. At our facility, we have sequenced around 75 different organisms comprising of more than 427 genomes and 730 transcriptomes - prokaryotes and eukaryotes (fungi, plants and animals). In addition we have optimized other unique applications such as small RNA (miRNA, siRNA etc), long Mate-pair sequencing (2 to 20 Kb), Coding sequences (Exome), Methylome (ChIP-Seq), Restriction Mapping (RAD-Seq), HLA typing, Genotyping By Sequencing (GBS), Mixed genomes (metagenomes) and Target Amplicons, etc. Translating DNA sequence data from NGS sequencer into meaningful information is an important exercise. Under NGGF, we have bioinformatics experts and high-end computing resources to dissect NGS data such as genome assembly and annotation, gene expression, target enrichment, SNP analysis, etc. Our services (sequencing and bioinformatics) have been utilized by more than 45 organizations (academia and industry) both within India and outside, resulting in many publications in peer-reviewed journals and several genomic/ transcriptomic data is available in NCBI/GenBank.

## Workflow at NGGF



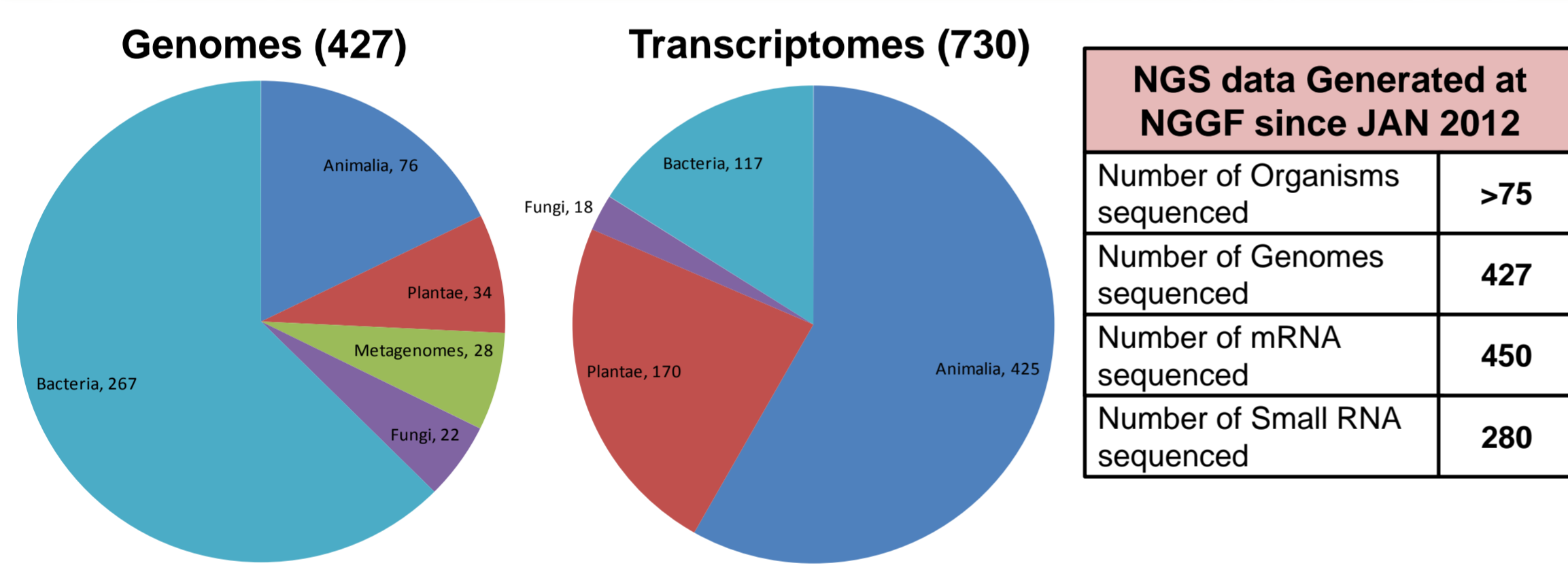
## Next-Generation Sequencing Platforms

Features	HiSeq (Illumina)	GS FLX + (Roche/454)
Sequencing Method	Sequencing by Synthesis	Pyrosequencing
Read lengths (bases)	100 x 2	Upto 1Kb
Data per run	Upto 300 Gb	Upto 1Gb
Run time	10 days	24 hours
Multiplexing (barcoding)	192 samples – 24 adapters & 8 lanes	192 samples – 12 adapters & 16 regions
Applications (examples)	Sequencing of <i>de novo</i> /reference genomes, transcriptomes (mRNA/small RNA), DNA/Chromatin modifications, SNP Markers etc.	<i>De novo</i> Sequencing of genomes, transcriptomes, HLA typing and metagenomes, targeted sequencing, etc.

## Bioinformatics and Data Analysis

Data Analysis Services	Expertise & Infrastructure
<ul style="list-style-type: none"> <li>Genome Assembly and Annotation</li> <li>Gene Prediction (<i>de novo</i> / Reference)</li> <li>Differential Expression Analysis</li> <li>Small RNA analysis</li> <li>SNP analysis</li> <li>ChIP – Seq analysis and other Protein – DNA/RNA interactions</li> <li>Metagenomics / Metatranscriptomics</li> <li>Genotyping by sequencing – RAD seq</li> </ul>	<ul style="list-style-type: none"> <li>Team of Qualified Analysts offering Personalized Services and Training</li> <li>Expertise to understand and solve problems using a systems biology approach</li> <li>High Performance Servers</li> <li>High Capacity Redundant Storage System with Optional Data Security</li> </ul>

## Organisms Sequenced at NGGF



Kingdom	Examples of Organisms Sequenced (genomes/ transcriptomes)
Animals	Mouse, Human, Rat, Antelope, Fish, Drosophila, Silkworm, Helicoverpa, White Fly, Nematode, Snail, Planaria, Hydra, etc.
Plants	Mango, Rice, Arabidopsis, Mangrove, Neem, Pitcher Plant, Tulsi, Foxtail Millet, Chilli, Brassica, Hordeum, Wheat, Drum stick, Methi, etc.
Fungi	Cercospora, Magnaporthe, Colletotrichum, Phytophthora, Alternaria, Ascochyta, Saccharomyces, Aschersonia, Paecilomyces, Beauveria, Aspergillus, etc.
Bacteria	Escherichia, Staphylococcus, Imtechella, Marinalabilia, Enterobacter, Vibrio, Serratia, Pseudoalteromonas, Alkalibacterium, Xenophilus, Caldimonas, Bacillus, etc.
Metagenomes	Human and Plant

## Major Projects

### Sequencing Projects

- Whole genome of *E. coli* - 114 genomes and 98 transcriptomes (NCBS).
- Transcriptome of Planaria - 84 RNA transcriptomes (inStem).
- Whole genome of microbes – 79 genomes (IMTECH).
- Sequencing of cancerous mouse: 36 genomes, 22 mRNA transcriptomes (Cambridge).
- Whole genome of fungal samples – 27 genomes (IIHR).
- Whole metagenomes – 24 human metagenomes (CMC).
- Rice Small RNA sequencing - 24 Small RNA transcriptomes (DRR).
- Cereal crop plants- 20 transcriptomes (NABI).
- Plant-Microbes genomes- Plants (Rice, Neem, Tulsi) and microbes (Bacillus) (C-CAMP).
- Plant transcriptomes – 50 transcriptomes (NIPGR).
- Insects – 24 transcriptomes (NCL).
- Medicinal Plants-18 transcriptome (NBRI).
- Whole genome of Plants- 5 genomes (CCMB).
- Transcriptome of Plants- 9 transcriptomes (NCBS).
- Transcriptome of Silkworm- 10 transcriptomes (CDFD).
- HLA typing - 80 samples (NCBS/BMST).
- RADseq - >100 samples (NCBS/UAS).
- 16s rRNA PCR Amplicon sequencing (NCBS).

### Data Analysis

- SNP analysis in mouse (Cambridge/inStem).
- ChIP seq analysis in Rat (JNCASR) and Chicken (IISER).
- miRNA analysis in Rice (DRR), Millet (NIPGR) and Arabidopsis (NIPGR).
- Assembly and annotation in Staphylococcus (IISc).
- De novo* assembly and annotation in plant and endophytes genomes (C-CAMP).
- De novo* assembly and annotation of Rice genome (CCMB).
- Allele mining for HLA loci - 80 samples (NCBS/BMST).
- SNP analysis Finger millet using RADseq data (UAS).
- Differential Expression Analysis in Human (inStem).

## Recent Publications

- Singh et al. *Grimontia indica* AK16T, sp. nov., Isolated from a Seawater Sample Reports the Presence of Pathogenic Genes Similar to *Vibrio* Genus. *PLoS One* 2014, 9(1):e85590.
- Sharma et al. Draft Genome Sequence of *Kocuria palustris* PEL. *Genome announcements* 2014, 2(1).
- Khatri et al. Genome sequencing and annotation of *Amycolatopsis azurea* DSM 43854. *Genomics Data* 2014, 2:44-45.
- Sasidharan et al. miRNAs in the planarian *Schmidtea mediterranea*. *RNA*. 2013 Aug 23.
- Krishna et al. 2013. small RNA repertoire in *Hydra magnipapillata*. *Nucleic Acids Res*, 2013 Jan 7;41(1):599-616.
- Prabhakara et al. 2013. Genome of *Staphylococcus aureus* ST772 from India. *PLoS One*. 2013; 8(3):e60013.
- Bala et al. 2013. Genome of *Rhodococcus ruber* Strain BKS 20-38. *Genome Announc*. March/April 2013 vol. 1 no. 2 e00139-13.
- Kaur et al. 2013. Genome of *Amycolatopsis decaplanina* Strain DSM 44594. *Genome Announc*. March/April 2013 vol. 1 no. 2 e00138-13.
- Kumar et al. 2013. Genome of *Rhodococcus triatomae* Strain BKS 15-14. *Genome Announc*. March/April 2013 vol. 1 no. 2 e00129-13.
- Kumar et al. 2013. Genome of *Streptomyces gancidicus* Strain BKS 13-15. *Genome Announc*. 2013 Apr 18;1(2):e0015013.
- Singh et al. 2013. Genome of *Acinetobacter baumannii* Strain MSP4-16. *Genome Announc*. 2013 Apr 4;1(2):e0013713.
- Srinivasan et al. 2013. Epistatic control of "expensive" genes in *Escherichia coli* K-12. *Mol Biosyst*. 2013 May 9.
- Bala M. et al. 2013. Draft Genome of *Rhodococcus qingshengii* Strain BKS 20-40. *Genome Announc*. 2013 Mar 28;1(2):e0012813.
- Prabhakara S. et al. 2012. Draft Genome of *Staphylococcus aureus* 118 (ST772), a major disease clone from India. *J Bacteriol*. 2012 Jul;194(14):3727-8.
- Khedkar S. et al. 2012. Draft Genome of *Staphylococcus aureus* ST672, an emerging disease clone from India. *J Bacteriol*. 2012 Dec;194(24):6946-7.
- Vikram S. et al. 2012. Draft Genome of Nitrophenol-Degrading Actinomyete *Rhodococcus imtechensis* RKJ300. *J Bacteriol*. July 2012; 194(13): 3543.
- Khatri I. et al. 2012. Draft Genome of *Rhodovulum* sp. Strain PH10 and *Alphaproteobacterium*. *J Bacteriol*. November 2012; 194(22): 6363.
- Kumar S. et al. 2012. Draft Genome of the marine bacterium *Marinilabilia salmonicolor* JCM 21150T. *J. Bacteriol*. 2012 Jul;194(14):3746.
- Kumar S. et al. 2012. Draft Genome of f the Halotolerant Bacterium *Imtechella halotolerans* K1. *J. Bacteriol*. July 2012;194(14) 3731.

Many more .....

## Summary

- Sequenced over 427 genomes (3.5 Tera bases) and 730 transcriptomes ( 4.8 Tera bases) ranging from prokaryotes to higher eukaryotes.
- Evolved as a National Facility to provide services to academic and industries within and outside India.

## Acknowledgement

We are thankful to Dept. of Biotechnology, Govt. of India for supporting the Next Generation Genomics Facility. We acknowledge our former facility members for their contribution to the facility: Ramya Malarini L, Manoj Kumar S, Anantharaman, Aarati Karaba, Annapurna S, Sivaraman K, Chellappa G and Jayanth B. We highly appreciate inStem and NCBS for supporting our facility.

## Core Facility Users

