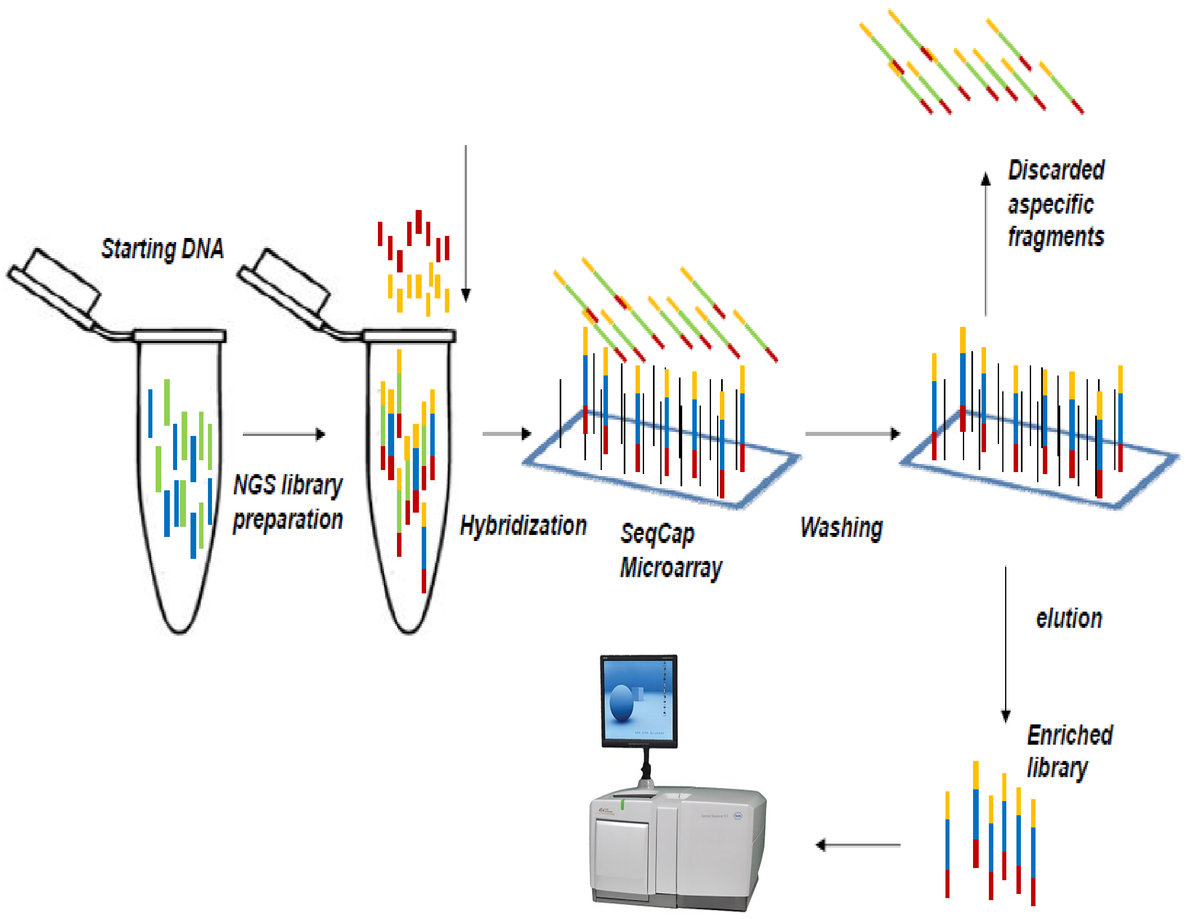
Sequence capture: a new tool for metagenomics

Pelletier Eric [1, 2, 3], Gyapay Gábor [1, 2, 3], Muselet Delphine [1, 2, 3], Richmond Todd 4, Le Paslier Denis [1, 2, 3]

1 Université d'Evry Val d'Essonne, Evry, France
2 CNRS, UMR8030, Génomique Métabolique, Evry, France
3 CEA, Institut de Génomique, Genoscope, Evry, France
4 Roche NimbleGen Inc., 500 South Rosa Road, Madison, WI 53719, U.S.A.

denis@genoscope.cns.fr

Introduction: Roche NimbleGen sequence capture technology was mainly used for the enrichment and then sequencing of the human exome (1,2). On the other hand, metagenomic analysis of complex environments usually leads to a high number of small contigs and a high proportion of uncompleted genes. For instance, this proportion was estimated to 40% for the human microbiomes projects (MetaHit and HMP). (3,4).



As a test case, we applied sequence capture to improve the quality of a gene set obtained from a complex metagenome: a full scale anaerobic digester from a waste water treatment plant (5)..

Initial material

Library	# reads
Fosmid	1,7 M
Plasmid	1 M
GS-FLX	342 K

Arachne initial assembly :

7495 contigs
2827 scaffolds
75.5 Mbp in total

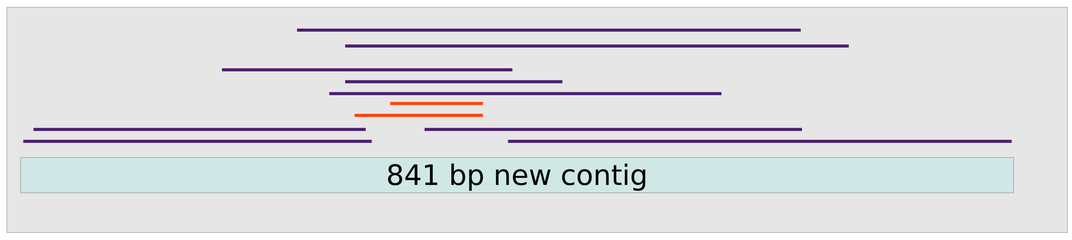
75,488 non redundant genes
10,785 being incomplete.

Two million probes were designed from uncompleted genes sequences (as well as from unassembled fosmid end sequences) and were used for genomic DNA capture and subsequent 454 Titanium sequencing.

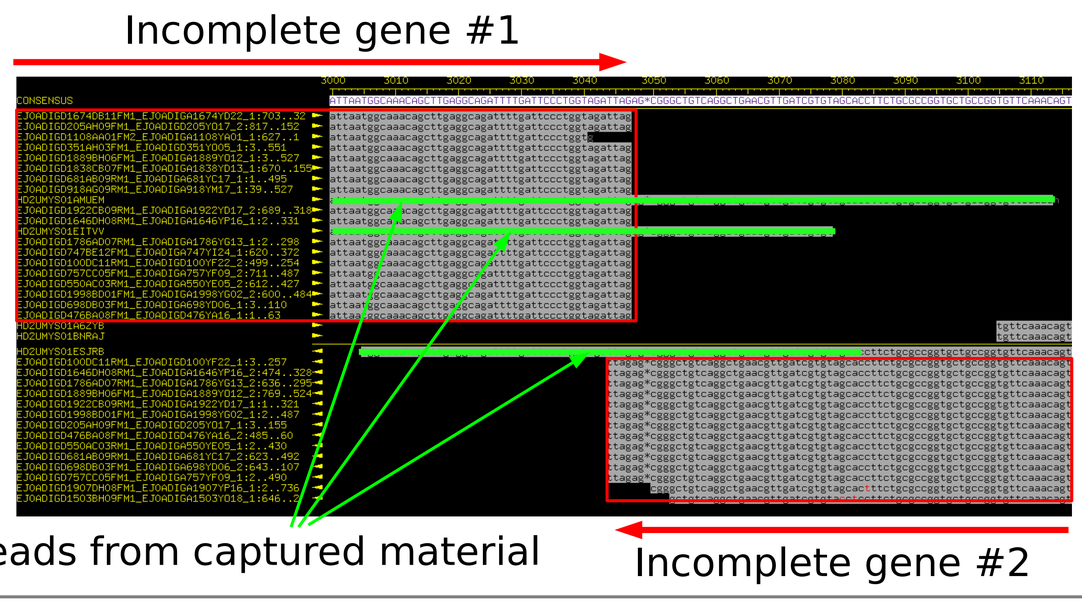
6,575 out of 10,785 previously uncompleted genes were completed, while 1236 other ones were extended.

Example 1 : extension of genes

2 FES initial Sanger reads with incomplete gene
9 454 reads from captured material



Example 2 : extension and merging of genes



In addition, Titanium run validated the original Arachne assembly and improved its quality by linking together a number of contigs and adding 25 Mb of new contig sequences.

In conclusion, sequence capture is an efficient technology, complementary to metagenomic analysis of complex environments. It allows to obtain better assemblies and to establish non-redundant complete genes catalogs. Moreover, due to the concentration and saturation properties of the capture mechanism, this approach allows to homogeneize the abundances of the target sequences from a complex initial population, giving a better access to low abundant material.
This technology could also be used for prokaryotic or eukaryotic genomic sequence finishing and re-sequencing genomes.

References
1 Basiardes S et al. (2005). Direct genomic selection. Nature Methods 1, 63-69.
2 Mamanova et al. (2010). Target-enrichment strategies for next generation sequencing, Nature methods 7:111-8.
3 Qin et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59-65.
4 Human microbiome project consortium. (2012). A framework for human microbiome research. Nature 486:215-21.
5 Pelletier et al. (2008). "*Candidatus* Cloacamonas acidaminovorans": genome sequence reconstruction provides a first glimpse of a new bacterial division. J Bacteriol 190:2572-79.