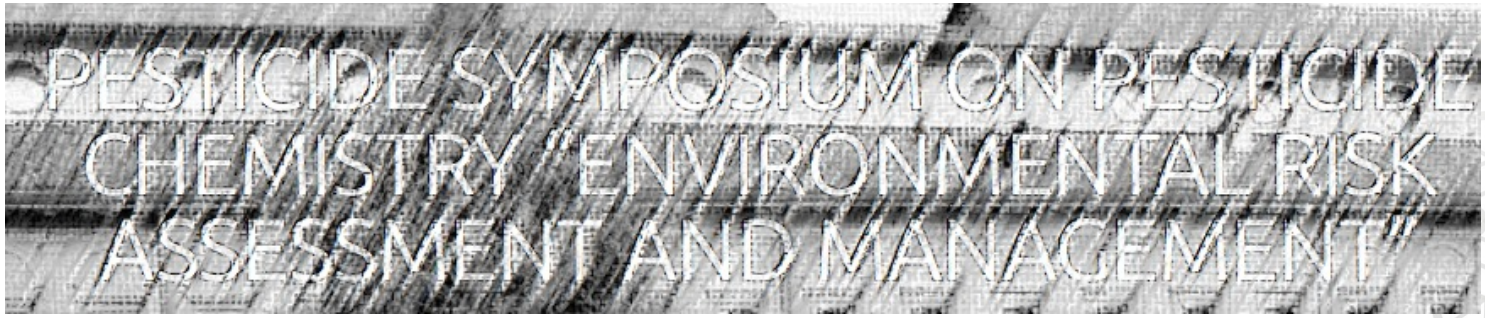


Development of a high density microarray for assessing functional diversity

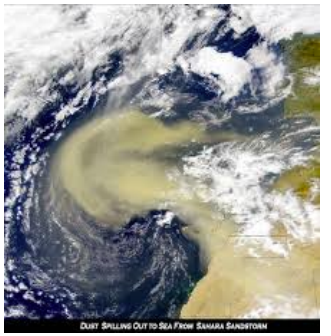
Molecular Genetics and Microbiology Group

Department of Environmental and Natural Resources Management
University of Patras, Agrinio, Greece

George Tsiamis

XV Symposium in Pesticide Chemistry

Potential functional diversity of atmospheric fine particles: How this can be assessed in a rapid, high throughput, and cost-effective way?

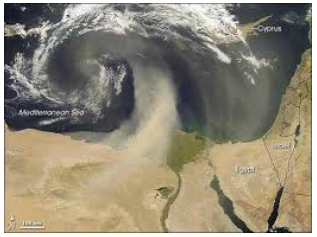


RT-qPCR – limited to a few genes per reaction

Advanced Molecular Approaches

Metagenomics
Metatranscriptomics

High density DNA microarrays
Functional gene arrays



Characterization of the functional gene diversity from atmospheric fine particles

LIMITATIONS OF A METAGENOMIC APPROACH

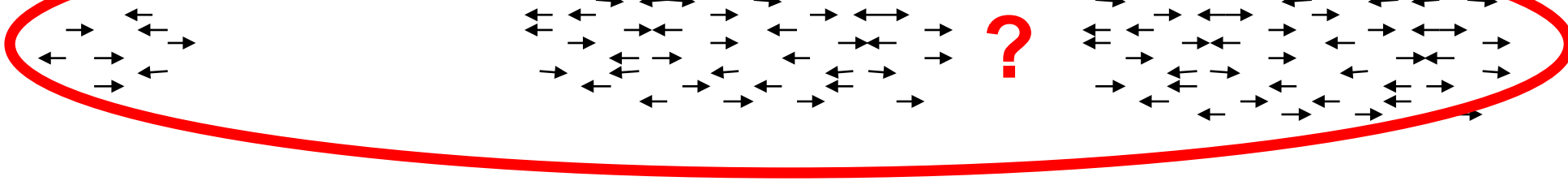
Acid mine drainage

Sargasso Sea

Soil

1 10 100 1000 10000

Species complexity



Microarray advantages

Because of their design DNA microarrays can provide information on microbial communities in a:

- simple
- rapid
- high-throughput and parallel manner
- semi-quantitative data
- most cost-effective than other molecular techniques
- more sensitive than other PCR based approaches
- modular, once a set of probes has been developed extra probes can be designed in order to increase the analysis of the microarray

Functional gene arrays

- Composed of probes for key genes involved in microbial functional process of interest
- FGA allow for the simultaneous examination of many functional genes, unlike PCR-based techniques that limit the number of genes that can be examined at one time

Table 1

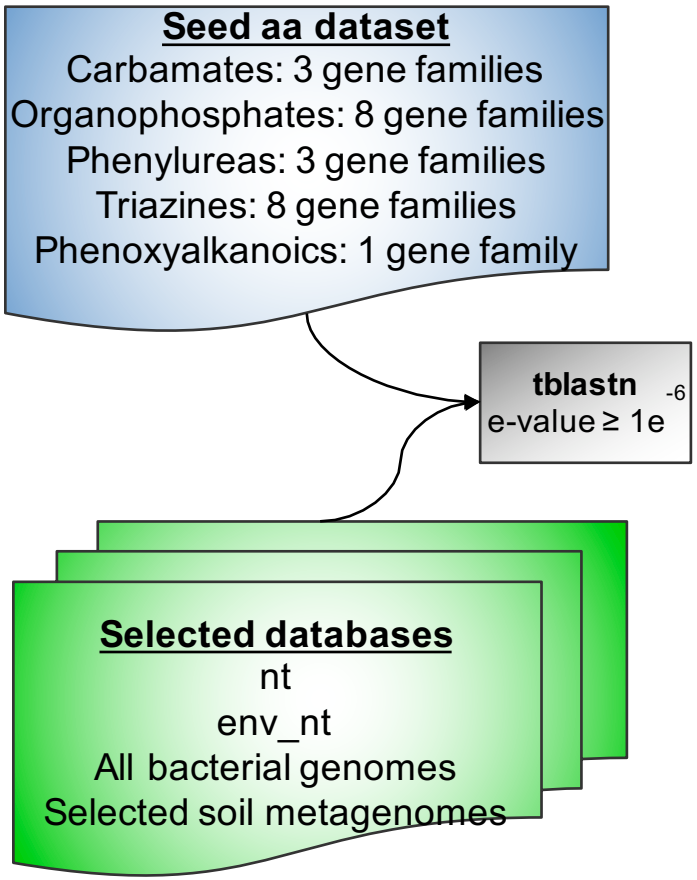
Summary of representatives of various functional gene arrays

Functional process (FGA)	No. functional gene families	Probe type	No. functional gene probes	Refs
N cycling and methanotroph	4 (amoA, nirS, nirK, pmoA)	Amplicons	89	[11]
N cycling	4 (amoA, nifH, nirS, nirK)	70-mer oligos	61 and 64	[25]
Antibiotic resistance	2 (tet, bla_{TEM-1})	Amplicons	18	[31]
N cycling (N fixation)	1 (*nifH*)	Amplicons	88	[26]
N cycling, methanotroph, S reduction	6 (*amoA*, *nirS*, *nirK*, *nifH*, *pmoA*, *dsrAB*)	50-mer oligos	763	[12]
Contaminant degradation, metal resistance	NS[a]	50-mer oligos	2042	[13]
N cycling (nodulation)	1 (*nodC*)	41–50 meroligos	130	[27]
Methanotroph	1 (*pmoA*)	~20-mer oligos	59 and 68	[29,30]
Virulence	2 (*invA*, *sopB*)	70-mer oligos	4	[32]
Virulence, antibiotic resistance	NS[a]	Amplicons	120	[33]
Comprehensive (GeoChip 2.0)	>150	50-mer oligos	24 243	[7]
Bioleaching	NS[a]	50-mer oligos	501	[37]
N fixation	1 (*nifH*)	~20-mer oligos	194	[28]
Virulence	>30	Oligos	791 and 2034	[35]
Virulence, antibiotic resistance (NimbleGen)	160	Oligos	1245	[34]
Antimicrobial resistance	NS[a]	Amplicons	800	[36]
Comprehensive(GeoChip 3.0)	292	50-mer oligos	27 812	[8••]
Comprehensive	NS[a]	cDNA clones	13 056	[38••]
Comprehensive (GeoChip 4.0, NimbleGen)	539[b]	50-mer oligos	120 054	[9•,10]

[a] NS: not specify.
[b] GeoChip also contains genes targeting human microbiomes with 36,062 probes in 139 functional gene families.

He at al., 2011 Current Opinion in Biotechnology

Pesticidechip – target sequence retrieval pipeline



General databases

nt

number of seq **29.628.407**
length **~85.2Gbp**

env_nt

number of seq **22.819.004**
length **~24.5Gbp**

bacterial genomes

number of genomes **25.304**
number of seq **3.237.238**
length **~98.3Gbp**

Soil metagenomic datasets

number of seq **69.573.255**
length **~24.5Gbp**

CAMERA

1. Waseca country farm soil metagenome
2. Metatranscriptomics of contaminated soil

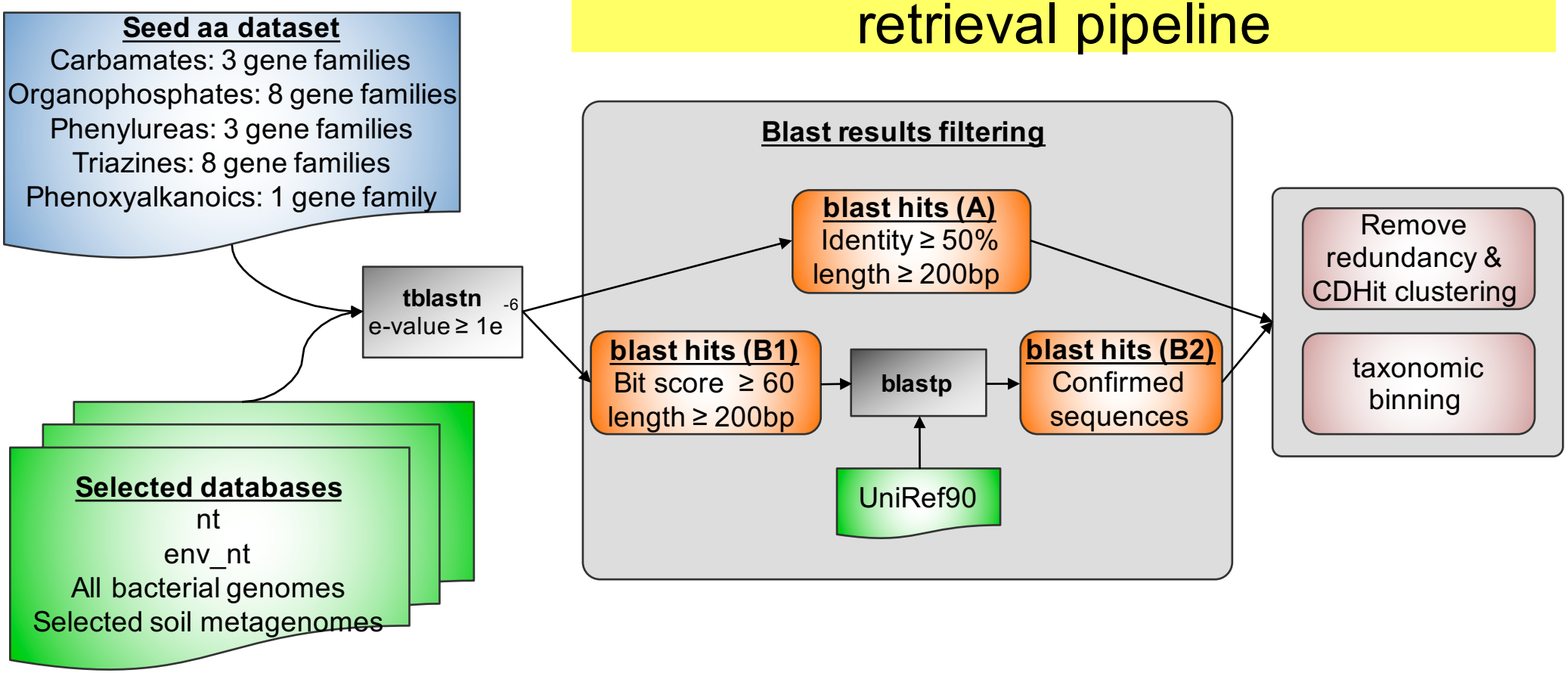
JGI

1. Soil microbial communities from Arlington Agricultural Research station (Project ID: 50510).
2. Soil microbial communities from Great Prairies (Project ID: 15780).

MG-RAST

1. Metagenomic analysis of HCH degrading soil microbial communities.
2. POME metagenome
3. Soybean Rhizosphere from Amazon soils

Pesticidechip – target sequence retrieval pipeline

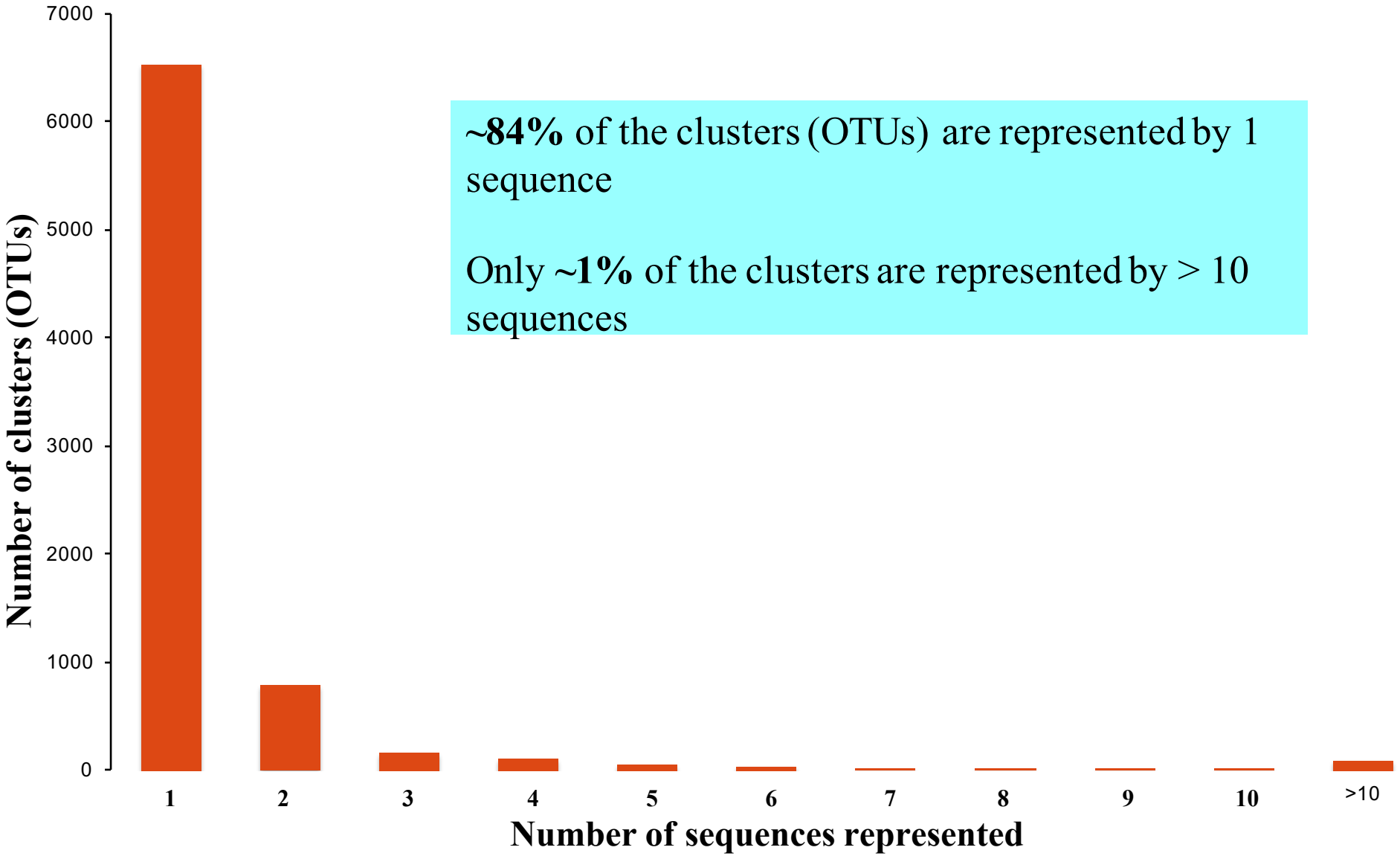


NOTES ON CREATING A SEED aa DATABASE

- **ophB** and **fedA** share **~97%** identities at aa level (*family 7*)
- **opdA** share **~67-87%** identities with **opd** group (*family 4*)
- **atzD** and **trzD** share >70% identities at aa level (family 16 and family 20 respectively) → merged (family16_20)
- The **mpd** group and the **ophC2** are distant homologues sharing **43-45%** identities at aa level (*family 5 and family 8 respectively*)
- **cahA** contains a conserved amidase domain (distant homology with Glutaminyl-tRNA synthase!)

Summary of sequence retrieval and clustering

(number of sequences per cluster - OTU)



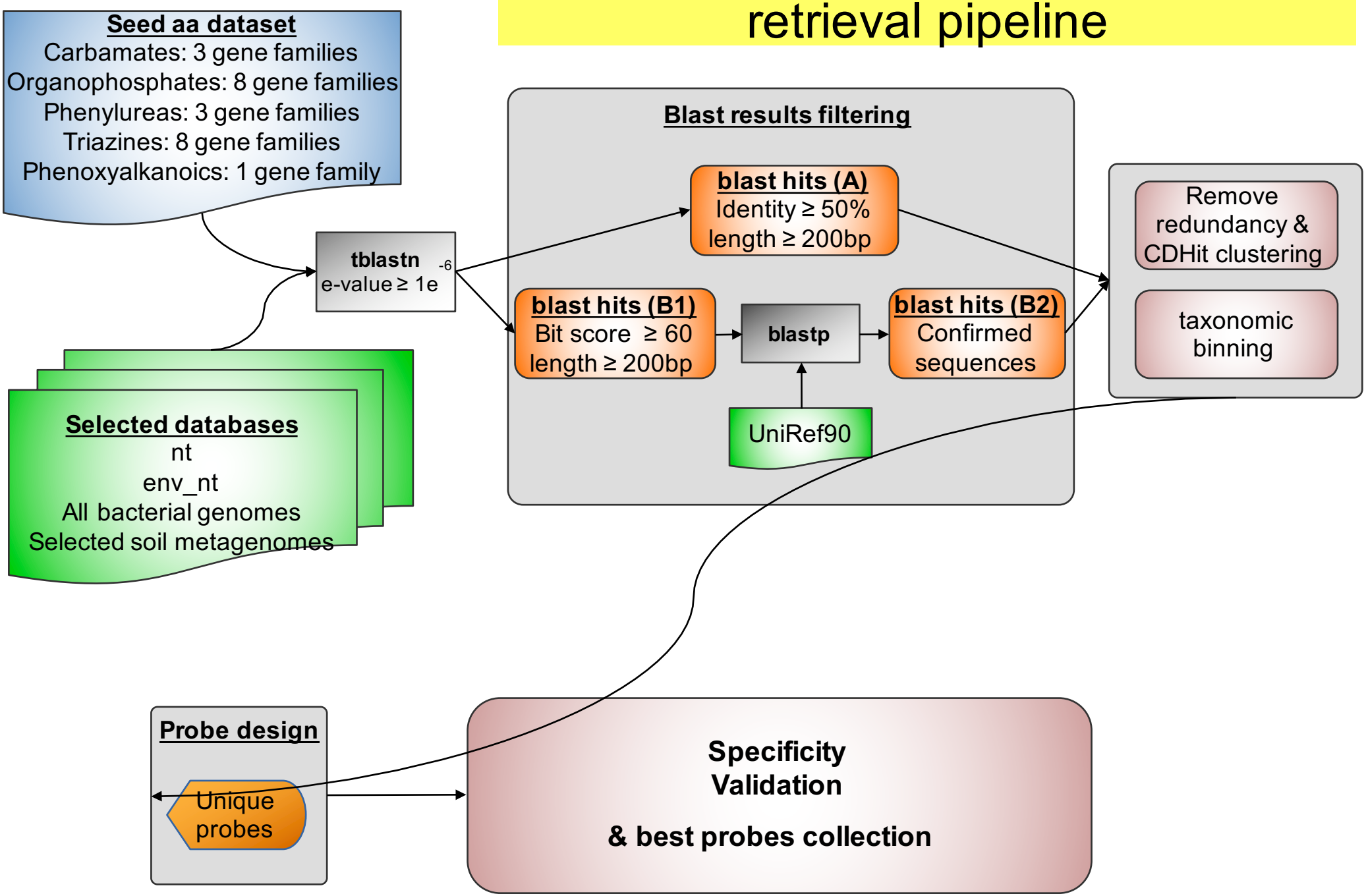
~84% of the clusters (OTUs) are represented by 1 sequence

Only **~1%** of the clusters are represented by > 10 sequences

Sequence divergence compared to seed sequences

gene family	mean	median	min	max
family1	53.99	52.835	36.28	100
family2	53.85	52.94	45.16	100
family3	72.31	98.83	34.86	100
family4	58.07	51.22	27.78	100
family5	64.48	56.25	39.23	100
family6	51.54	50.68	43.66	100
family7	48.03	50.27	37	100
family8	56.73	55.37	31.62	100
family9	78.67	84.58	27.4	100
family10	55.23	53.54	44.32	100
family11	52.89	51.35	40.21	100
family12	54.2	51.48	44.02	100
family13	69.67	53.73	34.78	100
family14	63.06	57.94	50	100
family15	99.64	100	98.53	100
Family16_20	55.65	53.665	41.18	100
family17	54.23	53.72	43.61	100
family18	54.49	52.585	36.28	100
family19	83.73	98.25	35.87	100
family21	70.22	60.78	33.87	100

Pesticidechip – target sequence retrieval pipeline



Oligonucleotide probe design

ArrayOligoSelector

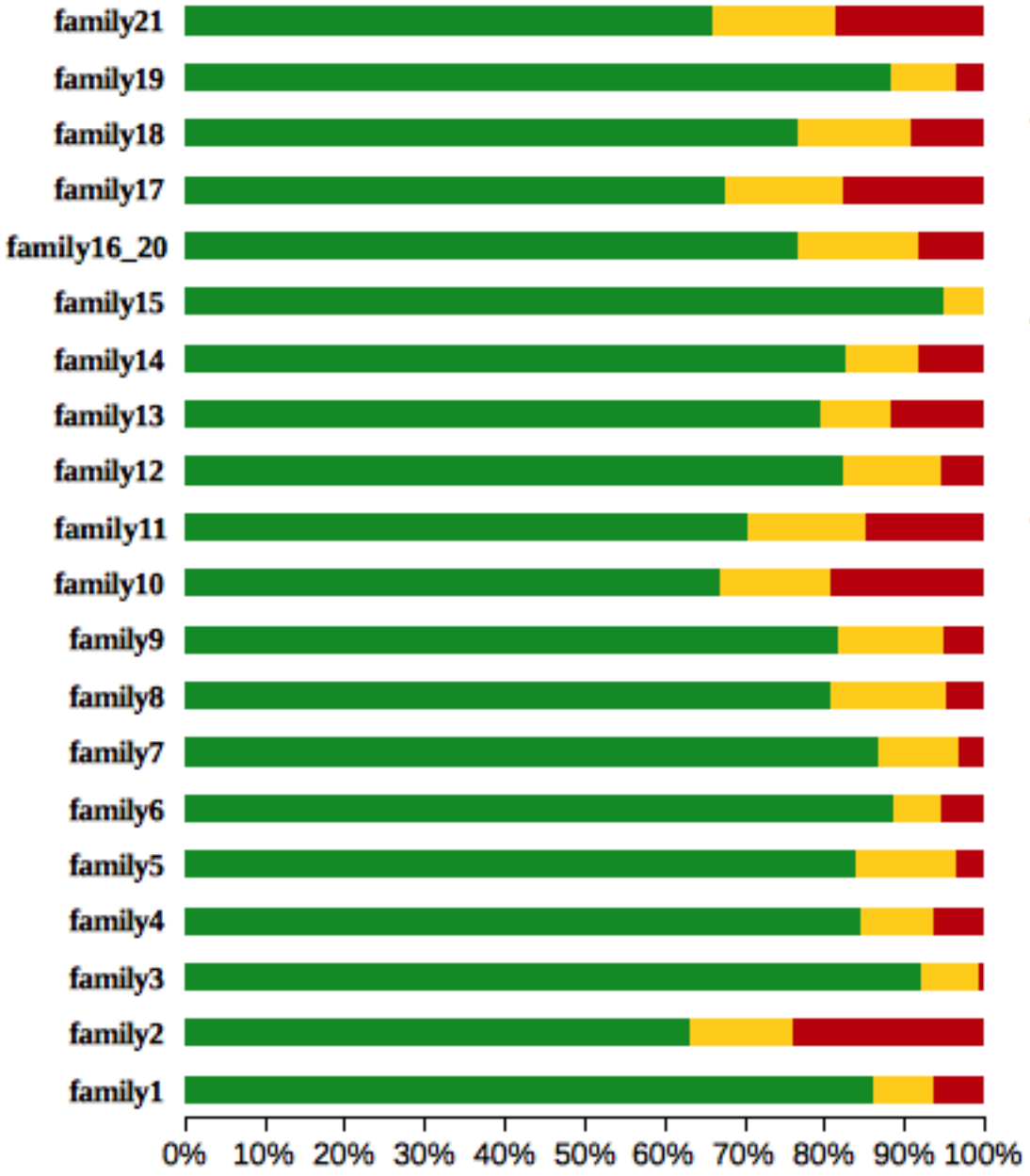
Cross-hybridization assessment based on Blast and thermodynamic calculations

Probe size = 50bp

Max number of probes designed by gene = 20

Non-target Blast database used for cross-hybridization assessment = complete bacterial genomes (genbank) and selected metagenomes

In-silico validation of oligonucleotide probes



- All probes were BLASTed against a non-target database (all metagenomes used without the target sequences)
- **Rejected:** probes with matches of ≥ 75% similarities and ≥ 95% coverage with non-target sequences
- **Marginal:** probes with perfect 20bp match in either 5' or 3' end with non target sequences
  - The binding strength of complementary ssDNA oligomers is altered by the presence of Inert Tails (*Di Michele et al.2014, J. Am. Chem. Soc.*)

Summary of oligonucleotide probes designed

Gene family	Probes designed (before validation)	good	marginal	bad
1	1501	1295	110	96
2	4874	3083	619	1172
3	129	119	9	1
4	595	504	53	38
5	1586	1330	199	57
6	2931	2603	173	155
7	465	404	46	15
8	1293	1047	185	61
9	2452	2006	320	126
10	1894	1271	258	365
11	384	271	56	57
12	402	331	49	22
13	147	117	13	17
14	336	278	30	28
15	20	19	1	0
16,20	927	711	141	75
17	7131	4820	1050	1261
18	7618	5844	1071	703
19	87	77	7	3
21	2220	1467	338	415
Total	**36992**	**27597**	**4728**	**4667**

CONCLUSIONS – FUTURE WORK

- Design of probes is completed
- The pipeline is semi-automatic with the development of new probes completed within 1-2 weeks per gene_family.
- 20 gene families, 8012 OTUs, 32,325 diagnostic probes
- DNA microarray will be developed as an open platform
- DNA microarray will be used in wet lab evaluation using reference soils (who is there=DNA; who is active=RNA)
- Development of a software package for the easy analysis of the microarray data
- Enrichment of the "functional microarray" with more gene families

Acknowledgements

➢ **University of Patras**

- Stefanos Siozios
- Sonia Nikolaki

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

chemisand
POPs / ELEMENTS /MICROBES | SAHARAN DUST

➢ **University of Crete**

- E. Stephanou
- M. Iakovidis
- S. Pergantis
- T. Tziaras

UNIVERSITY OF CRETE

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

ΕΣΠΑ 2007-2013
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Thank you!