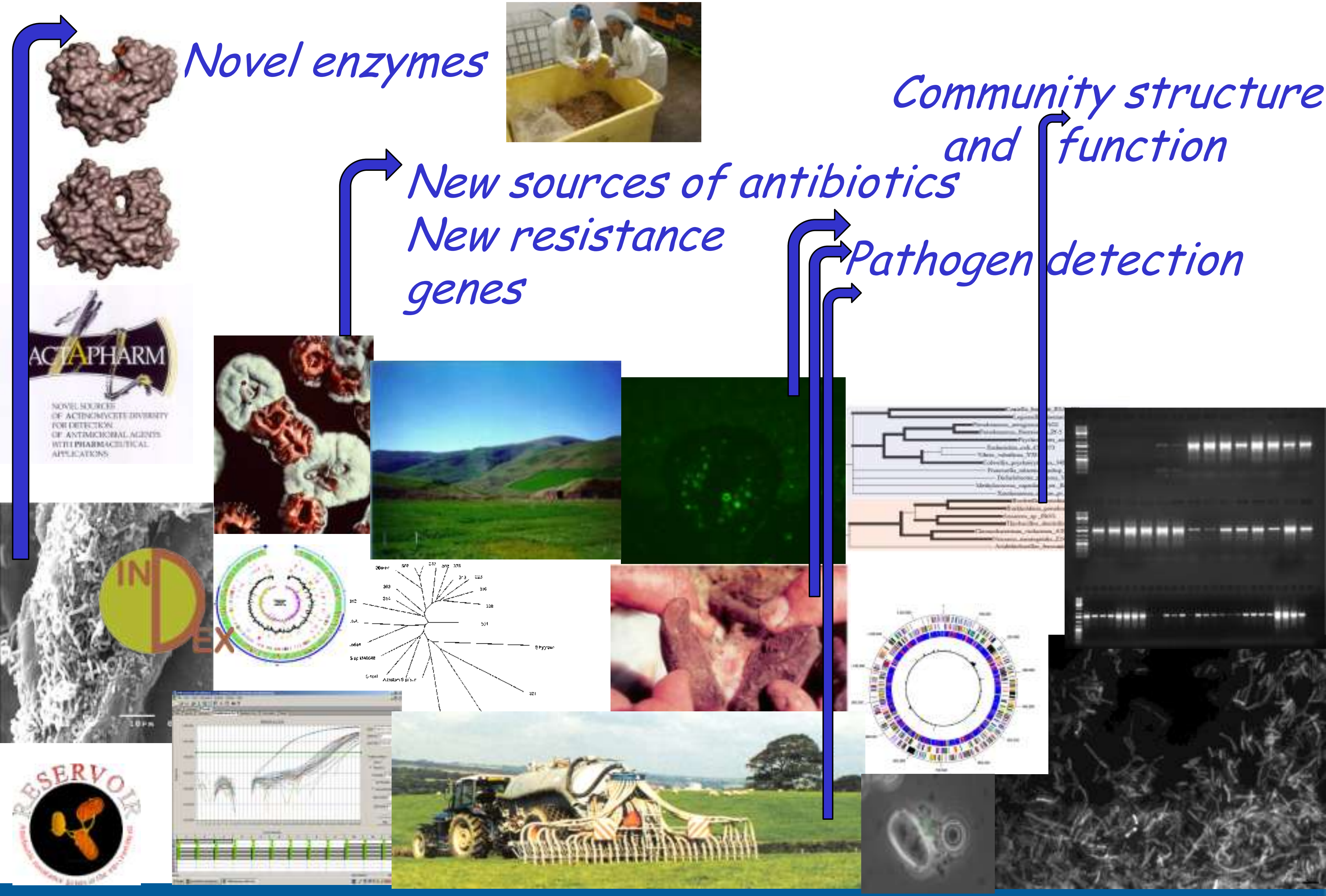
Microbial diversity, deep sequence analysis and metagenomes: the environmental gene pool



Novel enzymes

New sources of antibiotics

New resistance genes

Pathogen detection

Community structure and function

Metapopulation analysis

Metagenome:	-recovering uncultured bacteria -uncovering rare taxa -biogeography -capturing HMW DNA in expression libraries
Metatranscritome:	-IVET plant-microbe interactome -heavy metal resistome -antibiotic biosynthesis
Metaproteome:	-diverse proteins -enzymes after enrichments

Enzymes with useful properties for exploitation: some examples

1. Soil secretome: hydrolytic exoenzymes eg glycosyl hydrolases; chitinases
2. Soil antibiotic resistome: eg antibiotic disabling enzymes
3. Soil reservoir drug discovery: antibiotic biosynthesis enzymes eg polyketide synthases and non-ribosomal peptide synthetases

Family 19 chitinases

Endochitinases

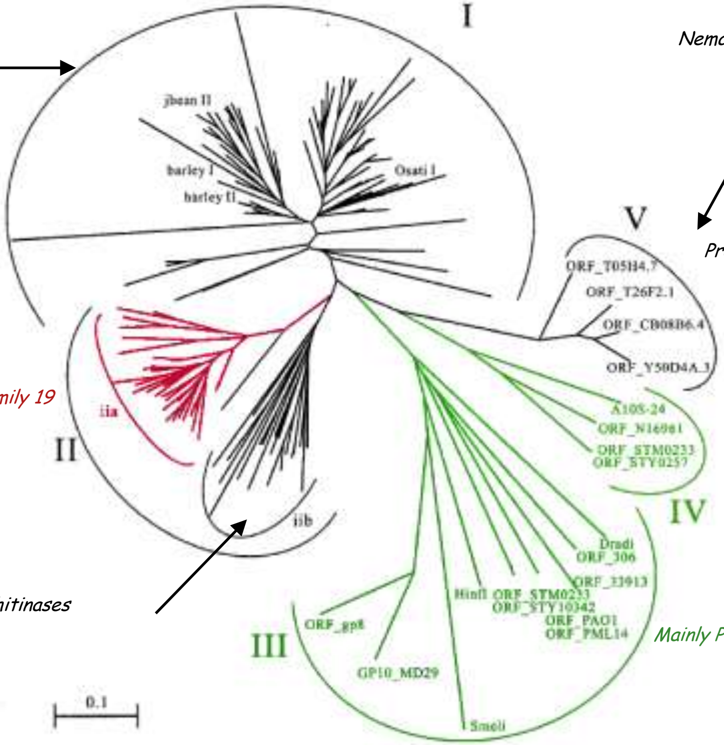
ENDOCHITINASES *have a groove-like structure which is open at both ends and cleave at random intervals within chain.*

Endo

EXOCHITINASES *have a tunnel-like structure and cleave chitooligosaccharides from the non-reducing end of a chitin chain.*

Exo

- *5 clusters*
- *Once thought only to occur in higher plants until found in Streptomyces griseus HUT 6037.* (Ohno, et al. 1996)
- *Actinobacteria dominate bacterial F19 databases*
- *Other bacteria are very distantly related*



Plant class I and II chitinases

Nematoda chitinases

Proteobacteria chitinases

Actinobacteria family 19 chitinases

Plant class IV chitinases

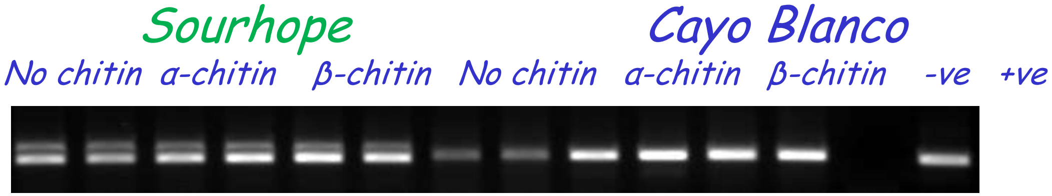
Mainly Proteobacteria and viral chitinases

(Kawase, et al. 2004)

Properties if family 18 and 19 chitinases

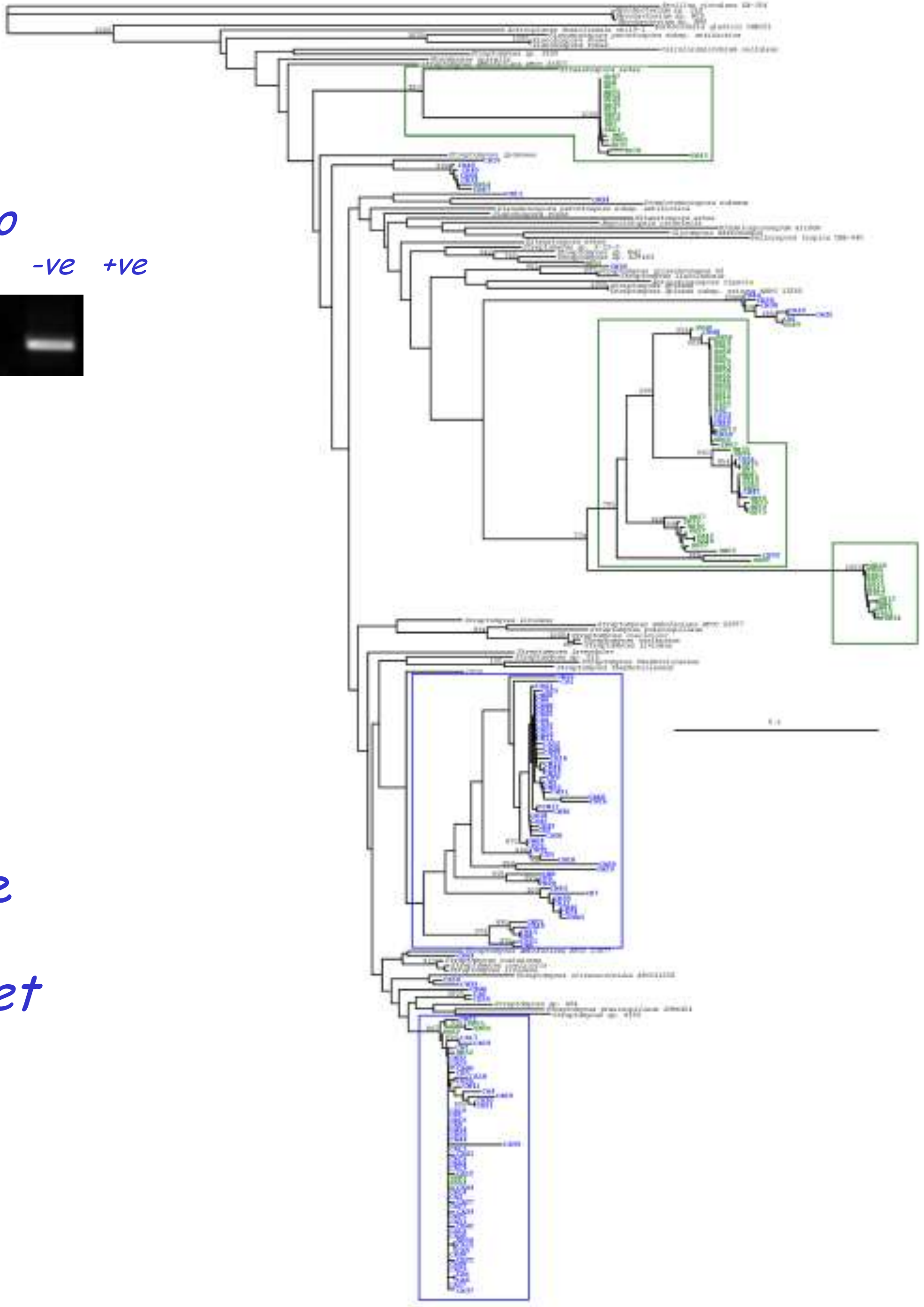
Multiplicity of genes Synergy of proteins	√ Family 18	? Family 19
Catalysis model	Substrate assisted	General acid-base
Mechanism	Retention mechanism	Inversion mechanism
Position of anomeric oxygen at C1	Equatorial (b)	Axial (a)
Exoxhitinase or Endochitinase	Exo- and Endo-	Endo-
Inhibitor (s)	Allosamidin	Amidines, amidrazones and nojiritetrazoles

Family 19 diversity and biogeography



PCR using family 19 chitinase primers

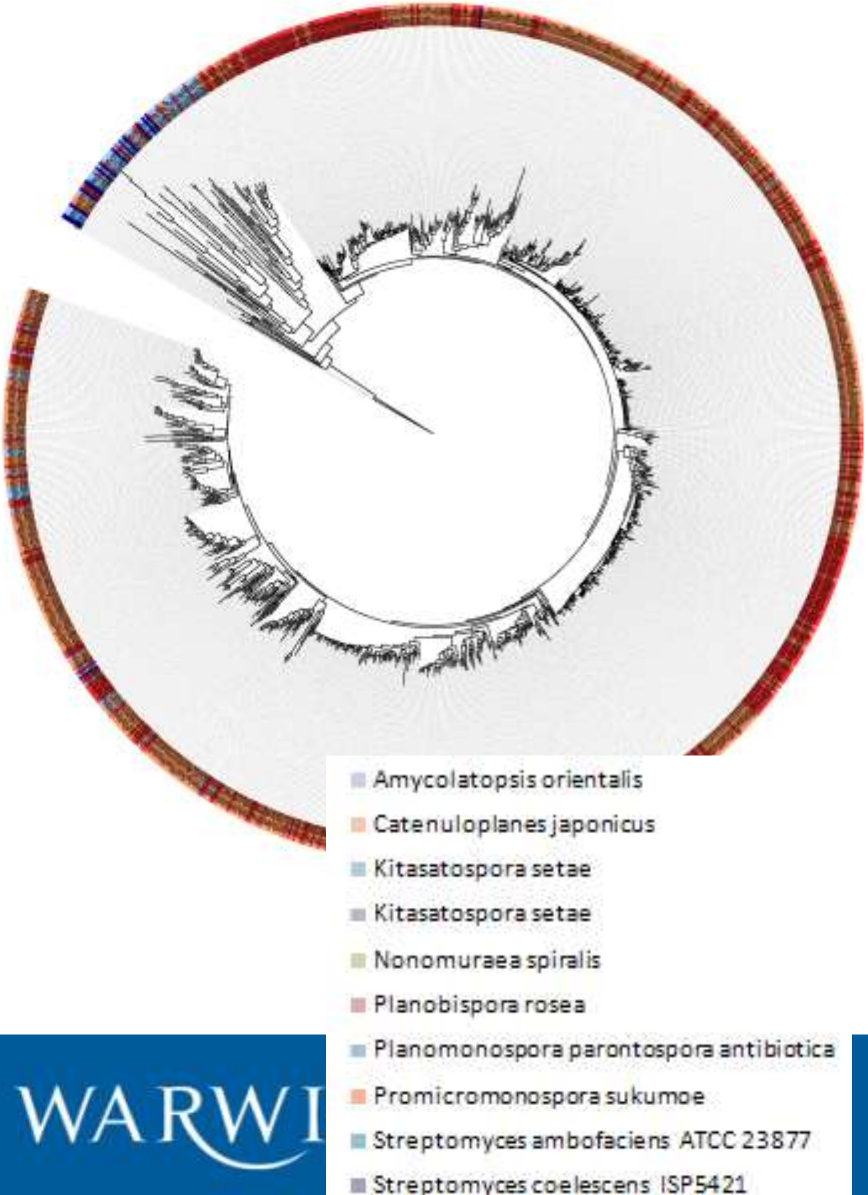
- Degenerate primers by (Williamson *et al.* 2000)
- Great diversity both within and between sites
- Family 19 chitinases may be specific to different allomorphs of chitin
- Physiochemical properties maybe more important than environmental factors (LeCleir, *et al.* 2004)

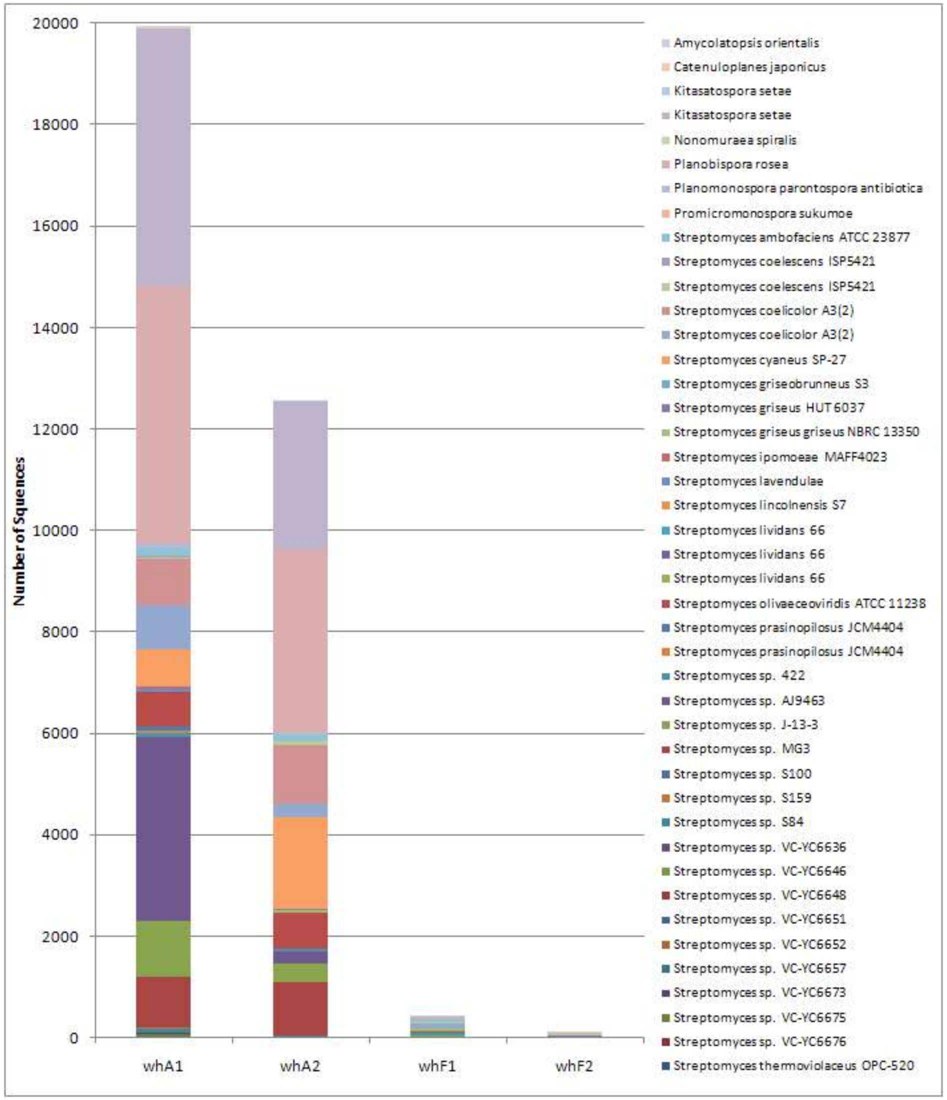


WARWICK

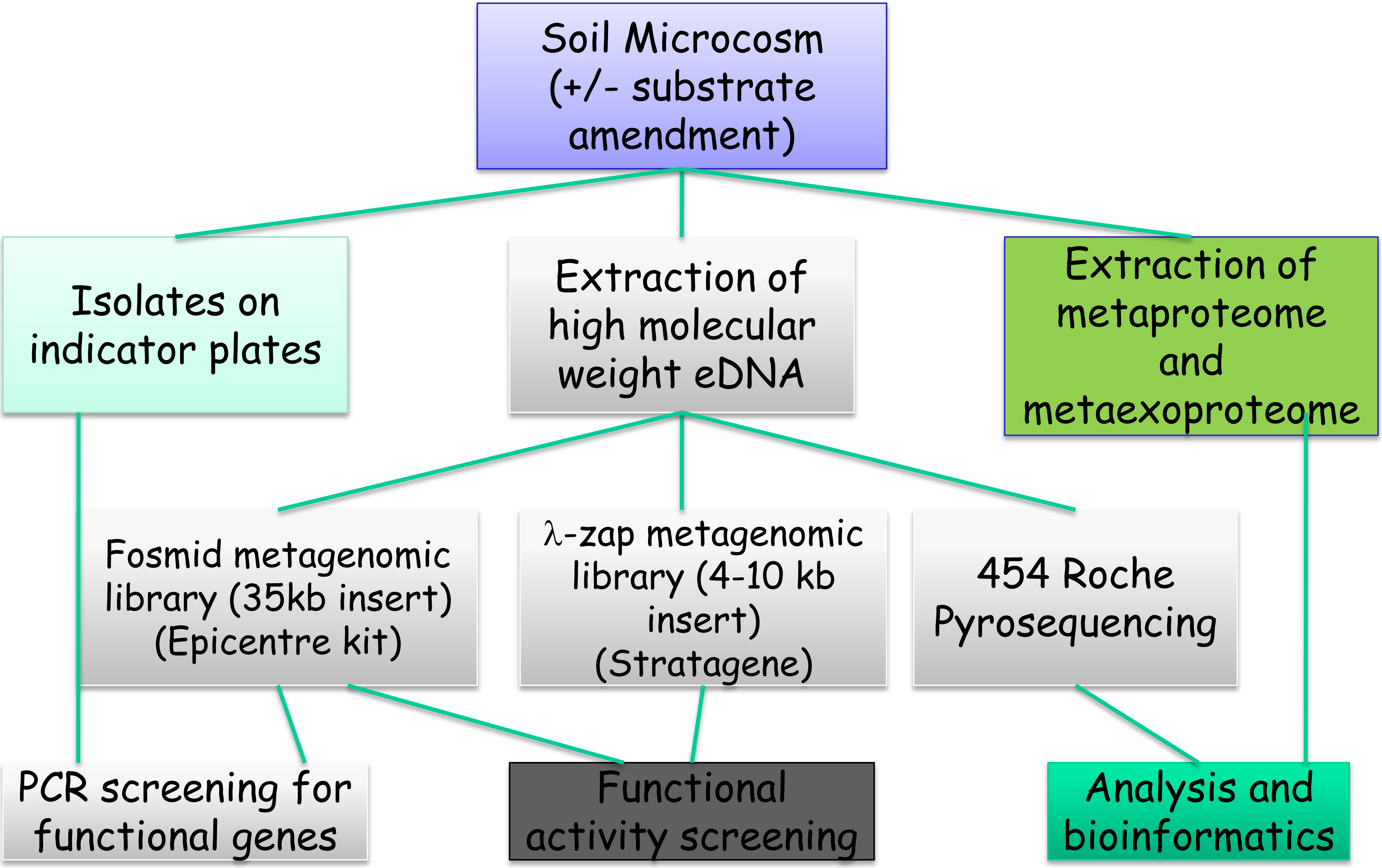
Enzyme molecular diversity

Enrichment alters diversity





Metagenomic and metaproteomic analysis of soil



Soil Microcosm (+/- substrate amendment)

Isolates on indicator plates

Extraction of high molecular weight eDNA

Extraction of metaproteome and metaexoproteome

Fosmid metagenomic library (35kb insert) (Epicentre kit)

λ-zap metagenomic library (4-10 kb insert) (Stratagene)

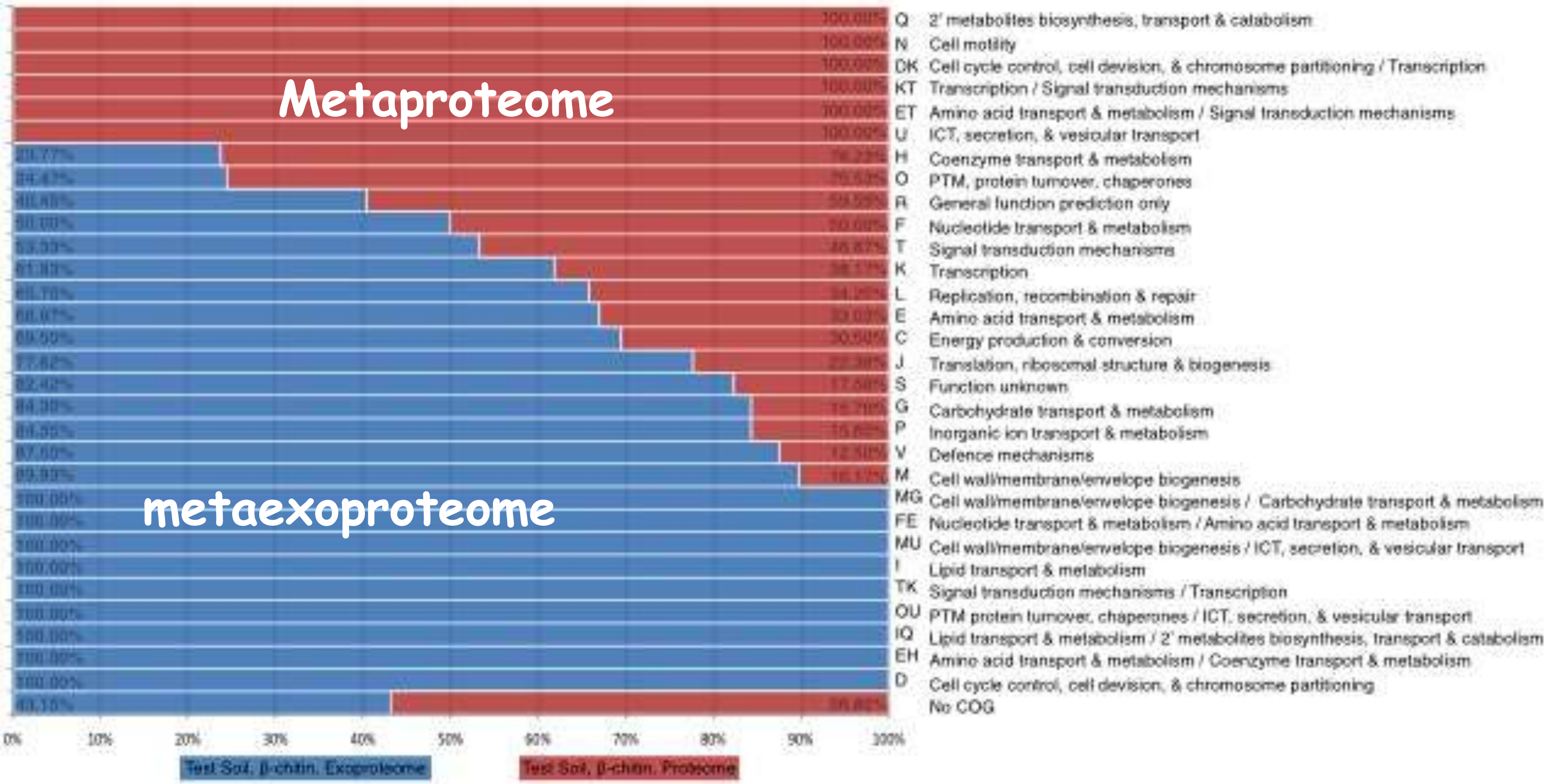
454 Roche Pyrosequencing

PCR screening for functional genes

Functional activity screening

Analysis and bioinformatics

Metaproteome combined with metaexoproteome of a soil

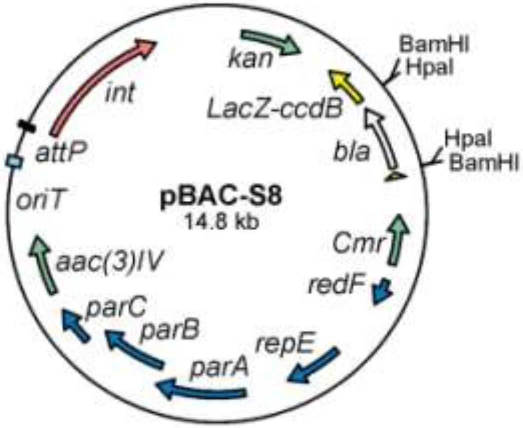


New sources of diversity from soil bacteria
Expression screening of metagenomic libraries

The **cellular biomass** was **collected** washed prior to **immobilisation** in an **LMP agarose ‘plug’**.

The **LMP ‘plug’** was **treated** with lysozyme, proteinase K and SDS to **lyse** the cells

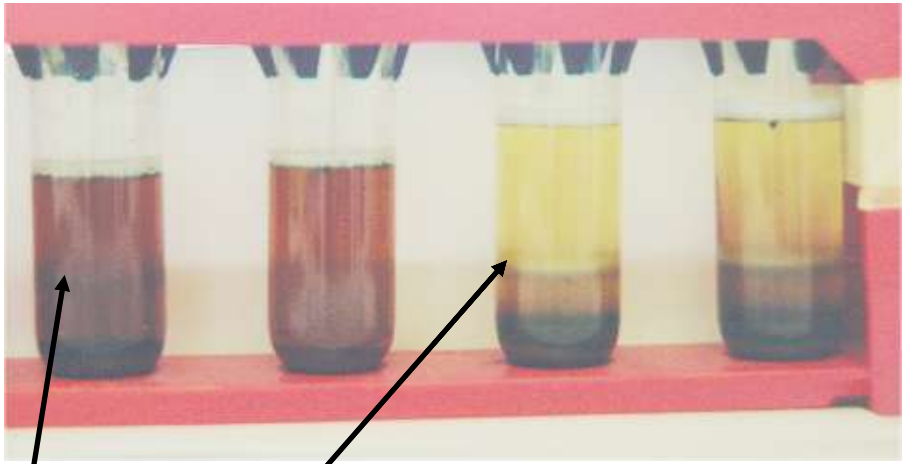
Bifunctional BAC vector for expression in *Streptomyces lividans*



(van der Geize et al., 2006)

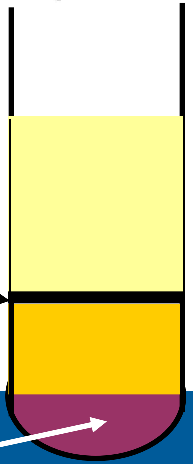
Air dried Warwick soil

Fresh Warwick Soil



Nycodenz extraction

(Bakken and Lindahl 1995)



Aqueous layer

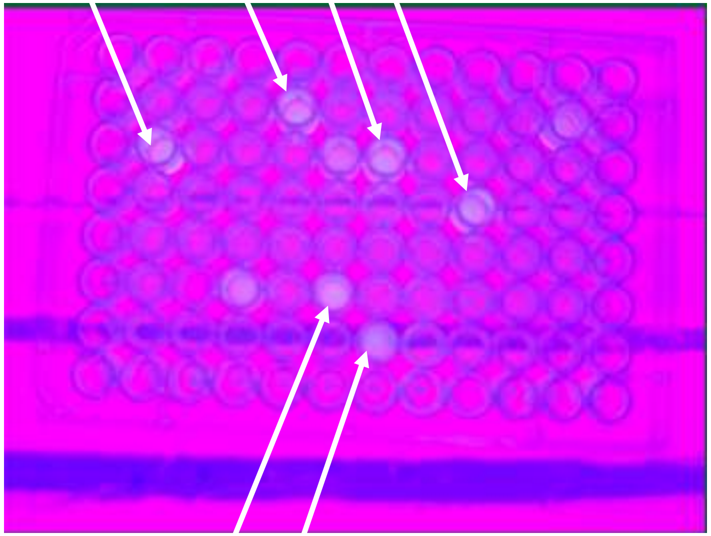
Nycodenz

Cell biomass at interface

Soil pellet

Metagenome libraries:

Expression screening for chitinase/lysozyme activity in lambda ZapII libraries



SOLR cells infected with phagemids to give c 150 clones per well. After 24 hrs of incubation 4-Muf-DiNag is added to 50mm.

The microtitre dish contains 14, 000 clones viewed under long-wave UV. Positive clones fluoresce as 4-MUF accumulates in the media.

Positive wells are diluted twice before plating to isolate single colonies.

The plasmid DNA has been isolated from +'s and transformed into the same backround to test for same positive signal. Inserts sequenced.

Screen improved by addition of F18 or F19 inhibitors and some of these inhibit lysozyme

Also screening fosmid and plasmid libraries inserts 40 → 4

Library sizes 0.5 → 2.6 Gb

Antibiotic resistance (soil resistome)

Soil acts as an environmental reservoir for antibiotic resistance genes –associated with ab biosynthesis clusters

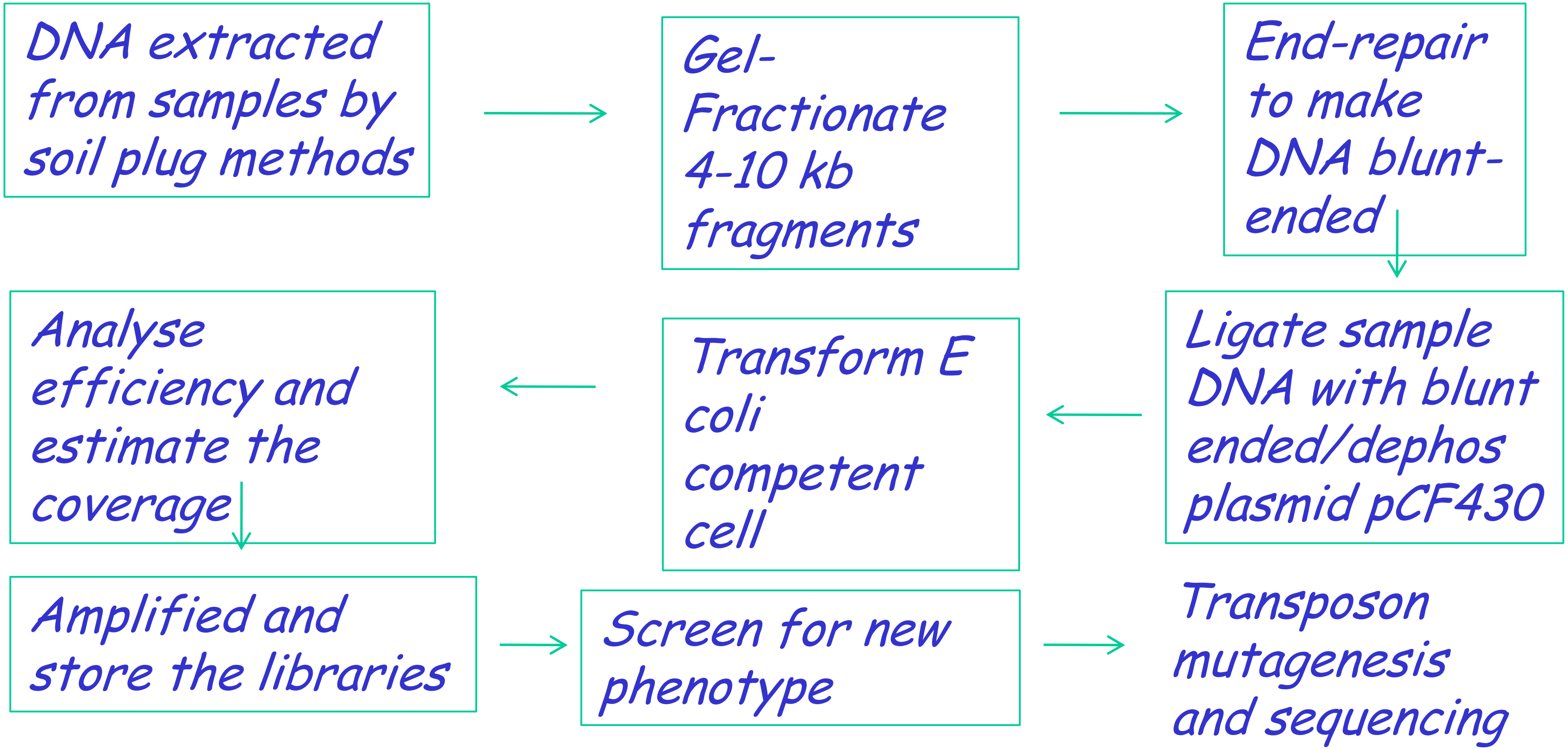
- -in closely related non-producers
- -in unrelated non-producers indigenous soil bacteria
- -in unrelated non-producers exotic bacteria = pathogens/commensals added to soil

- Potential for selection for resistance -pollution
- HGT of resistance genes- mobilome
- Pathogens can survive in soil
  - -Acquire integrons/plasmids
  - -Act as source of antibiotic resistance
  - – Possible carriage by amoebae-HGT

Reservoirs of antibiotic resistance genes in diverse environments: RESERVOIR survey

		Soil		Rhizosphere		Manure		Sewage		Seawater	
		Non-polluted	Polluted	Non-polluted	Polluted	Non-polluted	Polluted	Non-polluted	Polluted	Non-polluted	Polluted
Non-producers Streptomycin	*aph3*	●	●	●	●	●	●	●	●		
	aph6-Id	●	●	●	●	●	●	●	●		
	ant3	●	●	●	●	●	●	●	●	●	●
	adenylase			●	●	●	●				
	aph6-Ic	●	●	●	●						
	aph6-Ic (deg)	●	●	●	●	●		●		●	
Producers Streptomycin	***strA***	●	●	●	●	●	●	●	●		●
	aphD		●		●						
	strB1	●	●		●		●				
	stsC		●		●		●				
Non-producers Gentamicin	*aac(3)-I*	●	●	●	●	●	●	●	●	●	●
	aac(3)-II/VI	●	●			●	●	●	●		●
	aac(3)-III/IV					●	●				●
	aac(6')-II/Ib	●	●	●	●	●	●	●	●		●
	ant(2")-I	●				●	●	●	●	●	●
	aph(2")-I					●	●				
Non-producers Tetracycline	*tetA*		●			●	●	●	●		●
	tetB					●	●				
	tetC					●	●	●	●		●
	tetD					●	●				
	tetE		●		●	●	●	●	●		●
	tetG		●	●	●	●	●	●	●	●	
	tetH	●	●	●		●	●			●	●
	tetK	●	●	●	●	●	●				
	tetL		●		●	●	●	●	●		●
	tetM	●	●			●	●	●	●	●	●
	tetO		●			●	●	●	●		
	tetT										

Flow chart of metagenomic approaches



DNA extracted from samples by soil plug methods

Gel-Fractionate 4-10 kb fragments

End-repair to make DNA blunt-ended

Ligate sample DNA with blunt ended/dephos plasmid pCF430

Transform E coli competent cell

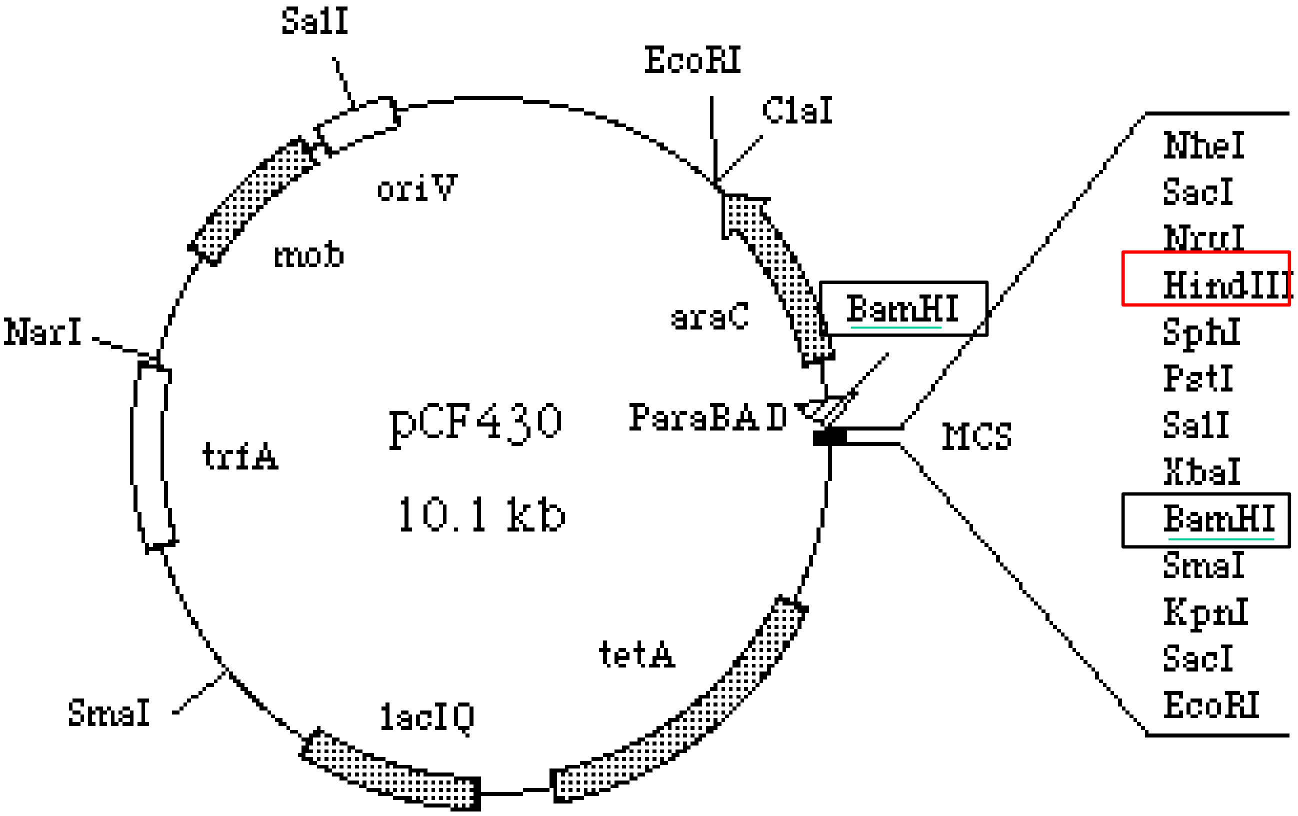
Analyse efficiency and estimate the coverage

Amplified and store the libraries

Screen for new phenotype

Transposon mutagenesis and sequencing

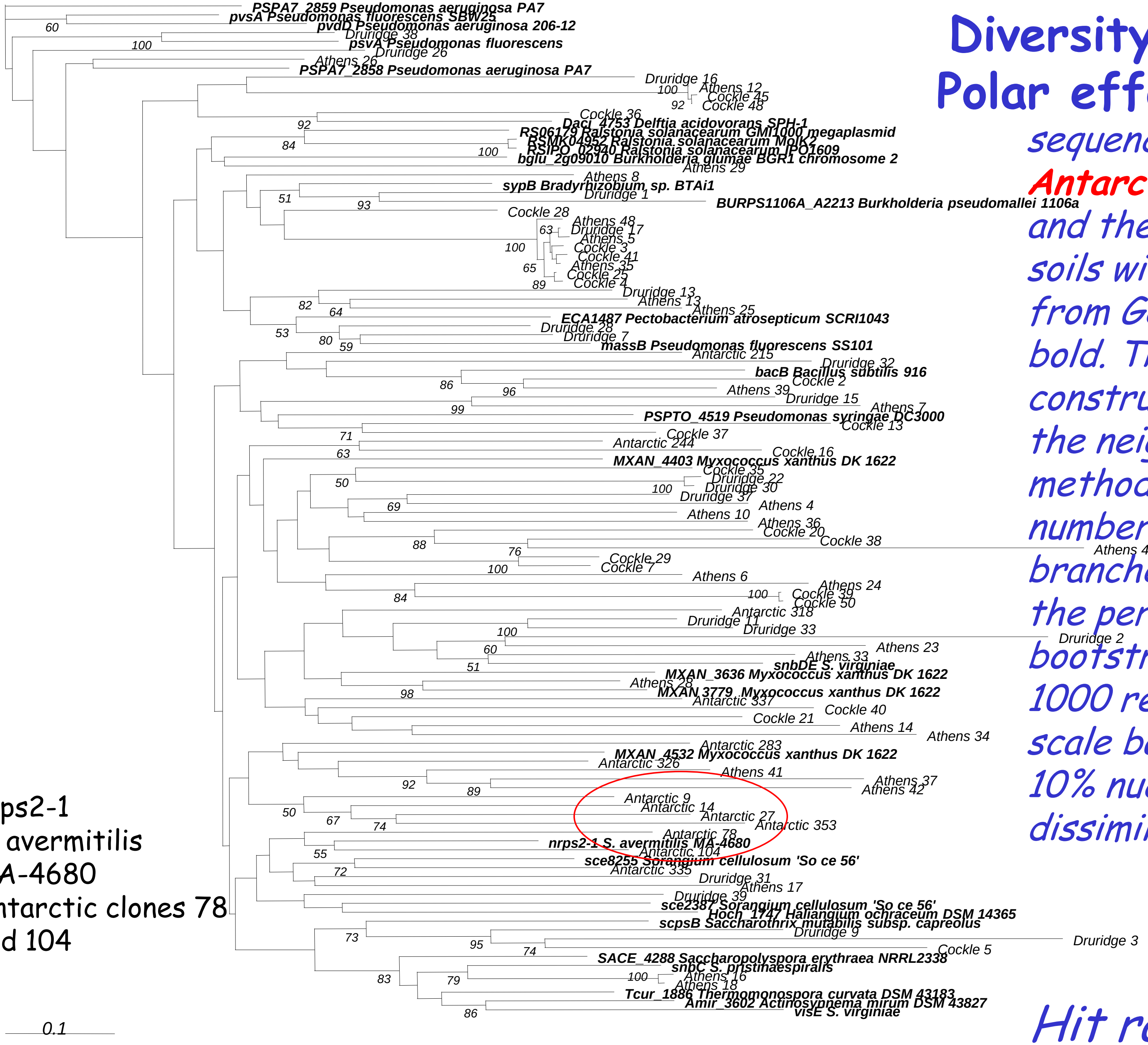
Broad-host-range expression vectors *that carry the -arabinose-inducible Escherichia coli araBAD promoter and the araC regulator*



Metagenomic Library and ESBL gene screening

	Reed Bed soil	Sewage Cake	1 Month Cake Applied	Control Soil	Grass Land Soil	FYM Applied Grass Soil
No. of clones		386 000	500000	170000	630000	210000
Average insert size (Kb)	4.64	4.12	4.40	3.70	2.85	3.71
Clones with inserts (%)	65	65	85	50	75	85
Coverage (Gb)	0.63	1.59	1.87	0.32	1.53	1.47
No of cefotaxime resistance	0	2	1	0	0	0
No of ceftazidime resistance	2	1	0	0	0	0
No of imipenem resistance	1 ?	2 ?	0	0	0	0
No of amp res clones	4 ?	5?	1 ?	0	0	0
Hit rate	1/80	1/150	1/900	0	0	0

Diversity of NRPS Polar effects?

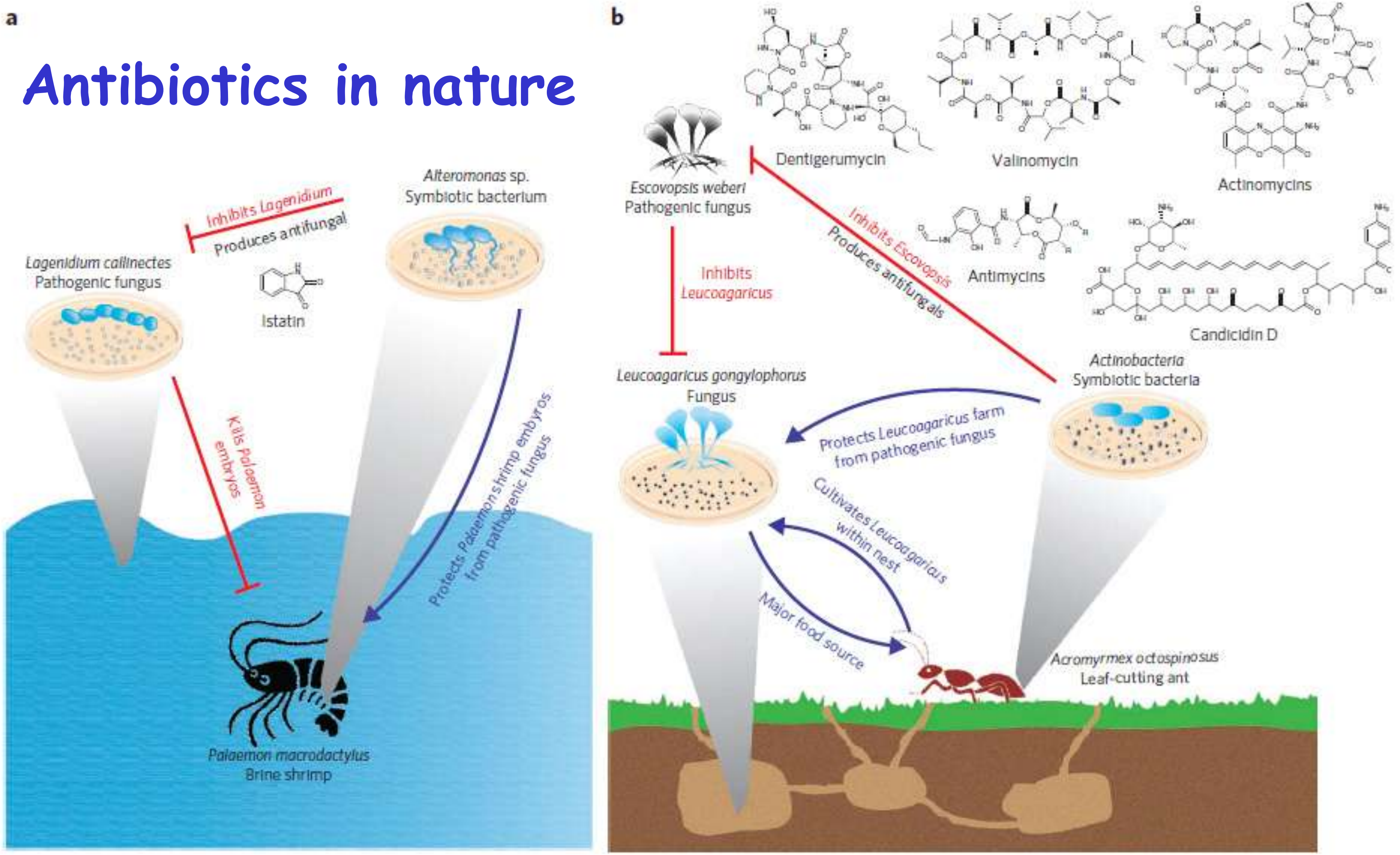


sequences from the ***Antarctic*** *library and the 3* ***European*** *soils with markers from GenBank in in bold. The tree was constructed using the neighbor-joining method; the numbers besides the branches indicate the percentage bootstrap value of 1000 replicates. The scale bar indicates 10% nucleotide dissimilarity*

nrps2-1
S. avermitilis
MA-4680
Antarctic clones 78
and 104

Hit rate 1/100

Antibiotics in nature



*Microbial metabolic exchange has important roles in ecology and the survival of higher organisms. (**a**) Symbiotic bacteria of the brine shrimp produce the antifungal compound istatin, thereby protecting shrimp embryos from pathogenic fungi. (**b**) Actinomyces spp. symbionts of leaf-cutting ants produce metabolites that protect the fungus farmed by the ants from a pathogenic fungus.*

Environmental mycobacteria as opportunistic pathogens

Particular risk to immunocompromised persons such as those with HIV/AIDS Pre-existing lung diseases and Helminth Infections

Clinical disease	Common causative species	Unusual causative species
Pulmonary Disease	*M. abscessus M. avium complex M. kansasii M. malmoense M. xenopi*	*M. asiaticum M. branderi M. celatum M. fortuitum M. gordonae M. haemophilum M. intermedium M. lentiflavum M. magdeburgensis M. shimodii M. simiae M. smegmatis M. szulgai*
Disseminated Disease	*M. avium complex M. chelonae M. haemophilum M. kansasii M. scrofulaceum*	*M. abscessus M. conspicuum M. fortuitum M. genavense M. malmoense M. marinum M. sherrisii M. simiae M. triplex M. xenopi*
Lymphadenitis	*M. avium complex M. malmoense M. scrofulaceum*	*M. abscessus M. bohemicum M. chelonae M. fortuitum M. haemophilum M. heidelbergense M. interjectum M. kansasii M. lentiflavum M. tusciae*
Cutaneous Disease	*M. abscessus M. chelonae M. fortuitum M. marinum M. ulcerans*	*M. haemophilum M.kansasii M. malmoense M. smegmatis*
Nosocomial Disease	*M. abscessus M. fortuitum M. chelonae*	*M. aurum M. avium M. gordonae M. mucogenicum M. neoaurum M. simiae M. smegmatis M. xenopi*

Atypical pulmonary disease

Cervical Lymphadenitis

Buruli Ulcer

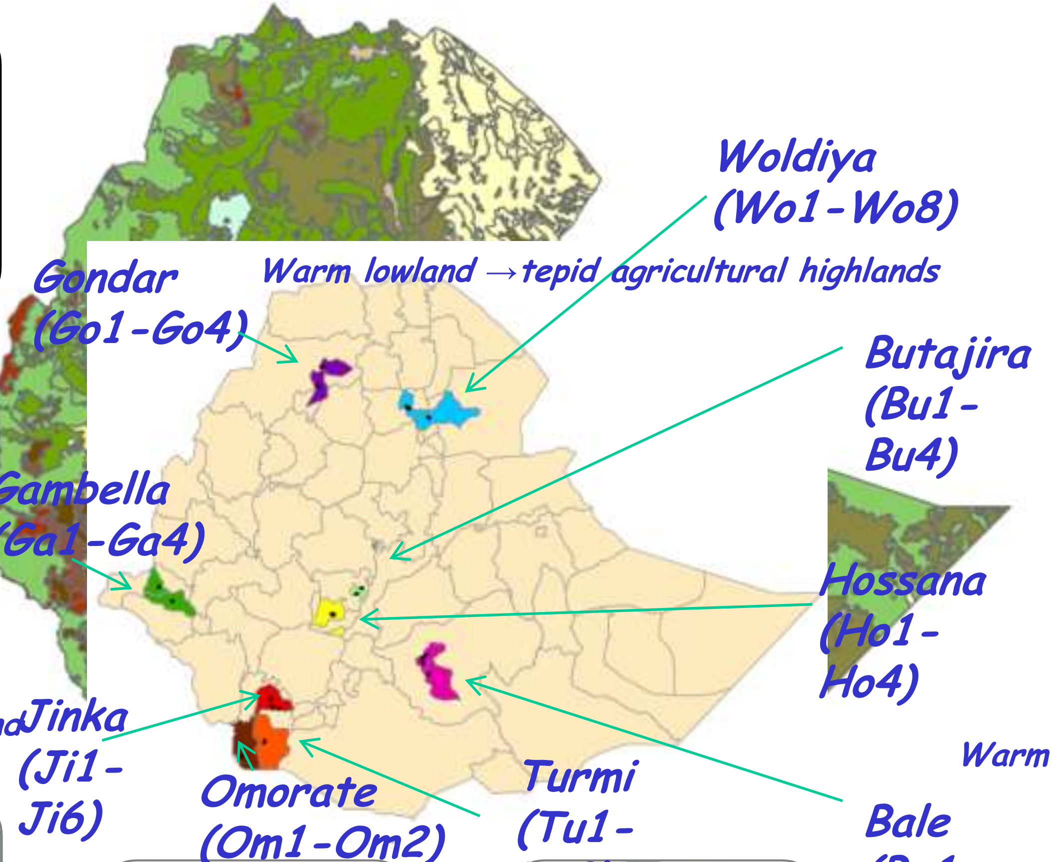
Aquarium Granuloma

Diverse sampling location: Ethiopia

Hot agricultural highland

Tropical agricultural lowland

Hot agricultural mid-highlands



Warm lowland → tepid agricultural highlands

Gondar (Go1-Go4)

Woldiya (Wo1-Wo8)

Gambella (Ga1-Ga4)

Butajira (Bu1-Bu4)

Hossana (Ho1-Ho4)

Jinka (Ji1-Ji6)

Omorate (Om1-Om2)

Turmi (Tu1-Tu2)

Bale (Ba1-Ba8)

Warm agricultural mid-highlands

Warm agricultural mid-highlands

Hot arid lowland

Hot arid lowland

Warm lowland → Dense forest → Afro-

Ethiopian EM diversity study

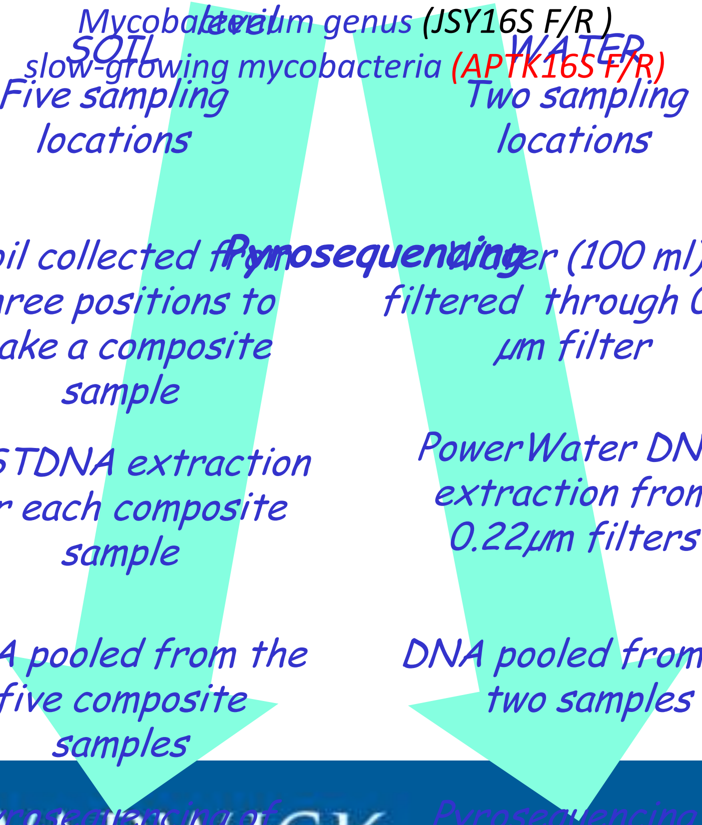
Culture independent 16SrRNA target gene

Two sets of primers to target EM

Mycobacterium genus (JSY16S F/R)

slow-growing mycobacteria (APTK16S F/R)

Sampling: village level



Pyrosequencing

SOIL

Five sampling locations

Soil collected from three positions to make a composite sample

FASTDNA extraction for each composite sample

DNA pooled from the five composite samples

Pyrosequencing of village soil DNA

WATER

Two sampling locations

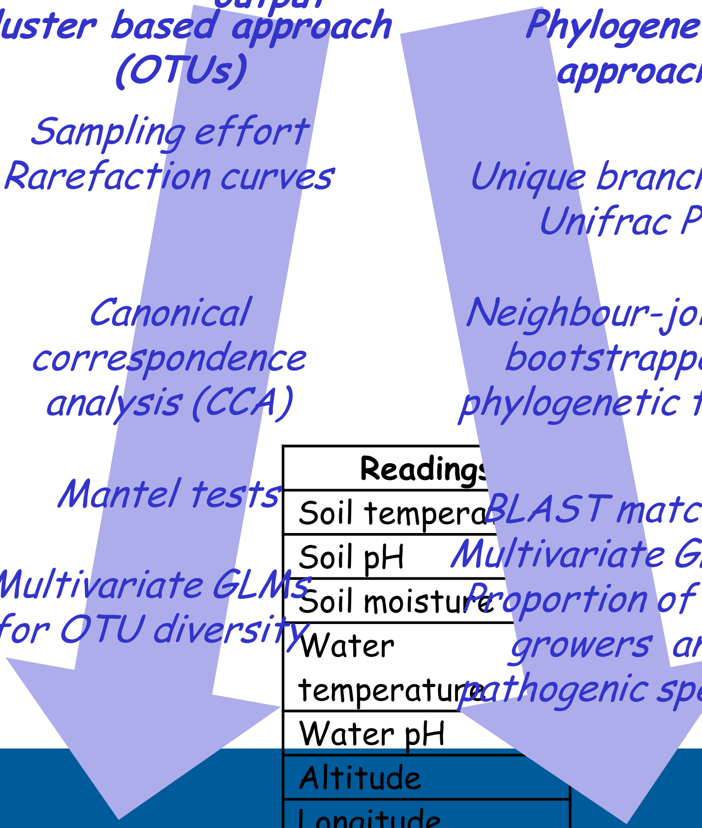
Water (100 ml) filtered through 0.22 µm filter

PowerWater DNA extraction from 0.22µm filters

DNA pooled from the two samples

Pyrosequencing of village water DNA

Pyrosequencing output



Cluster based approach (OTUs)

Sampling effort
Rarefaction curves

Canonical correspondence analysis (CCA)

Mantel tests

Multivariate GLMs for OTU diversity

Phylogenetic approach

Unique branch length
Unifrac PcoA

Neighbour-joining bootstrapped phylogenetic trees

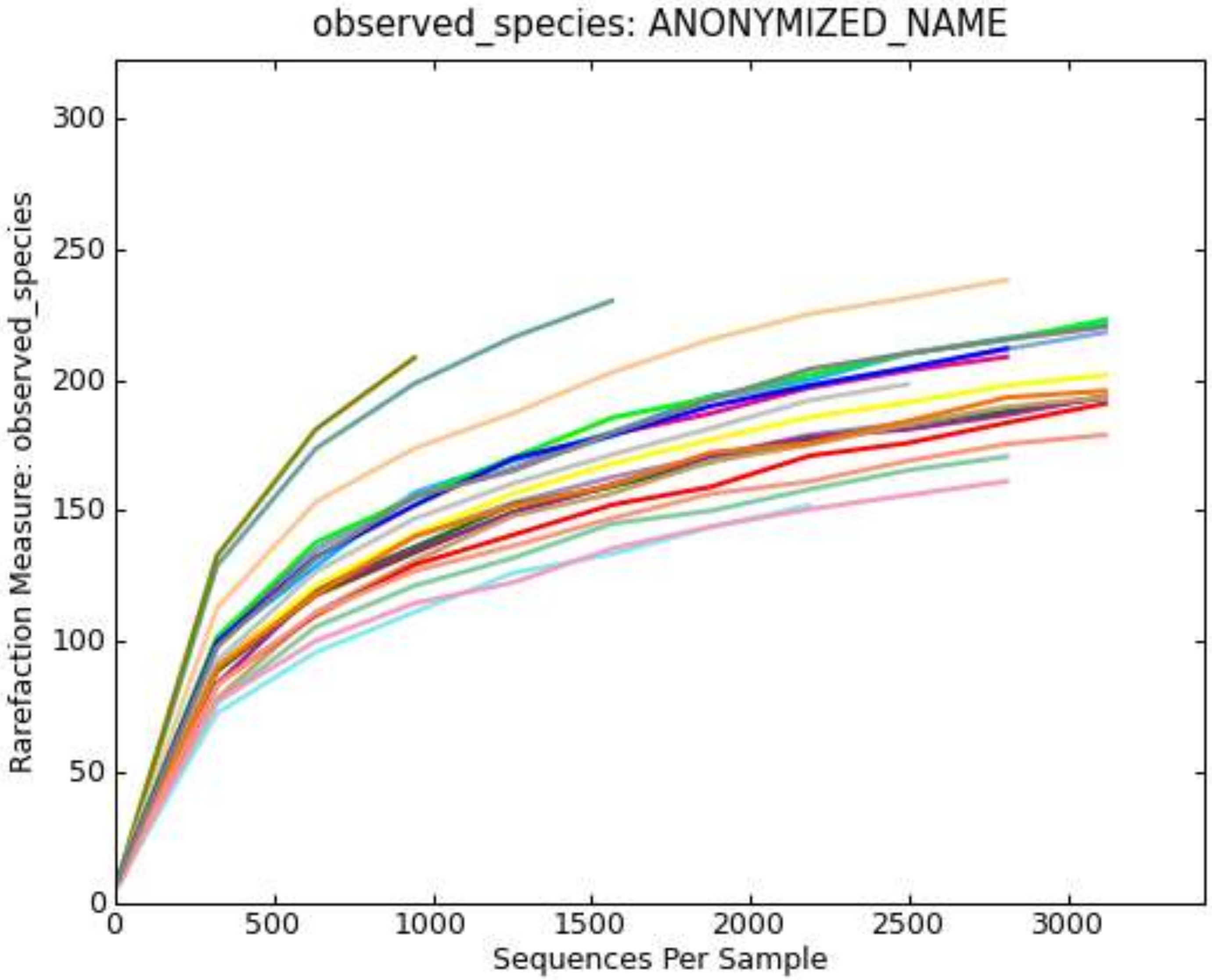
BLAST matches

Multivariate GLMs : Proportion of slow growers and pathogenic species

Readings
Soil tempera
Soil pH
Soil moisture
Water temperature
Water pH
Altitude
Longitude
Latitude

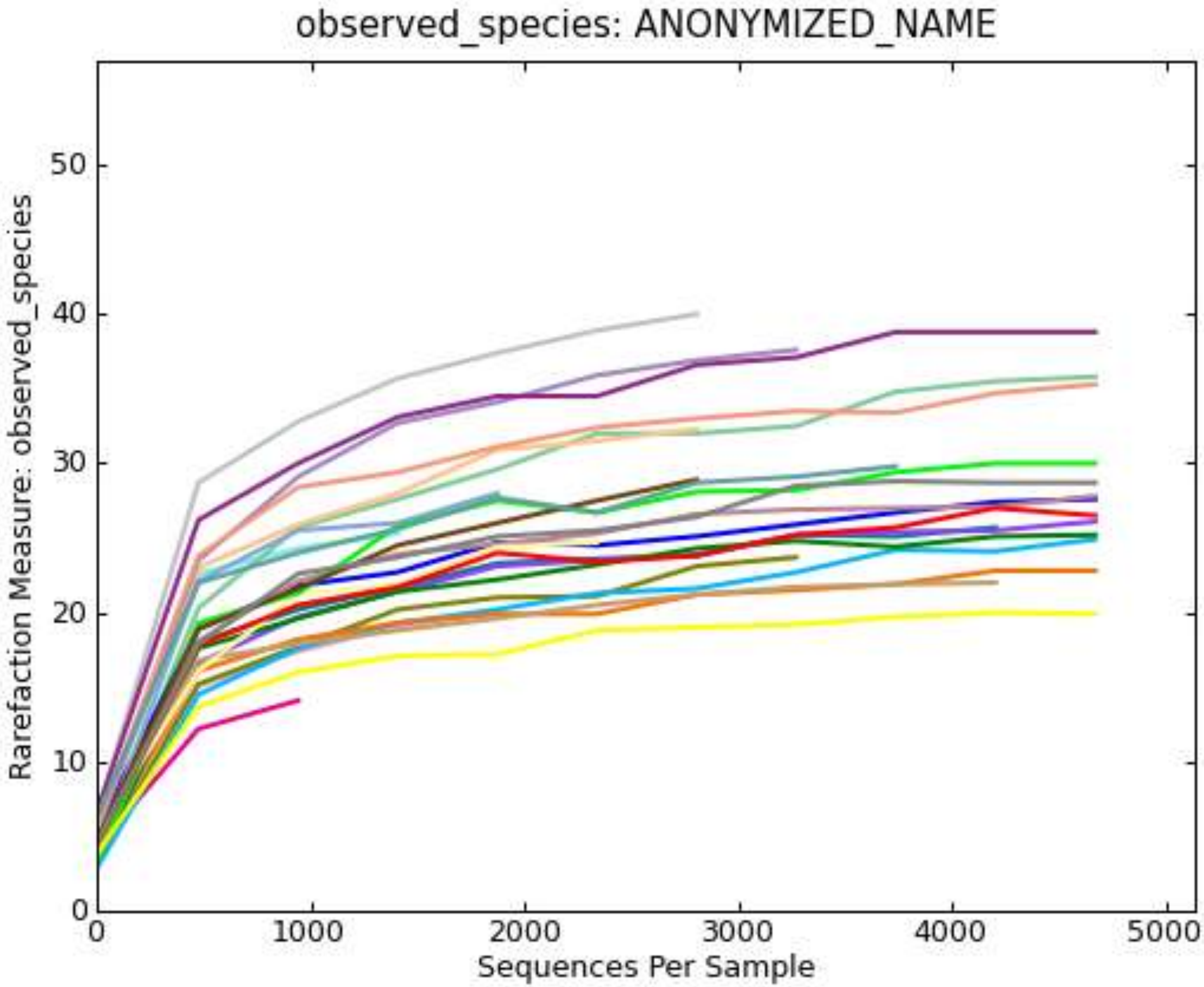
Mycobacterium genus level : DNA from soils

Alpha rarefaction curves



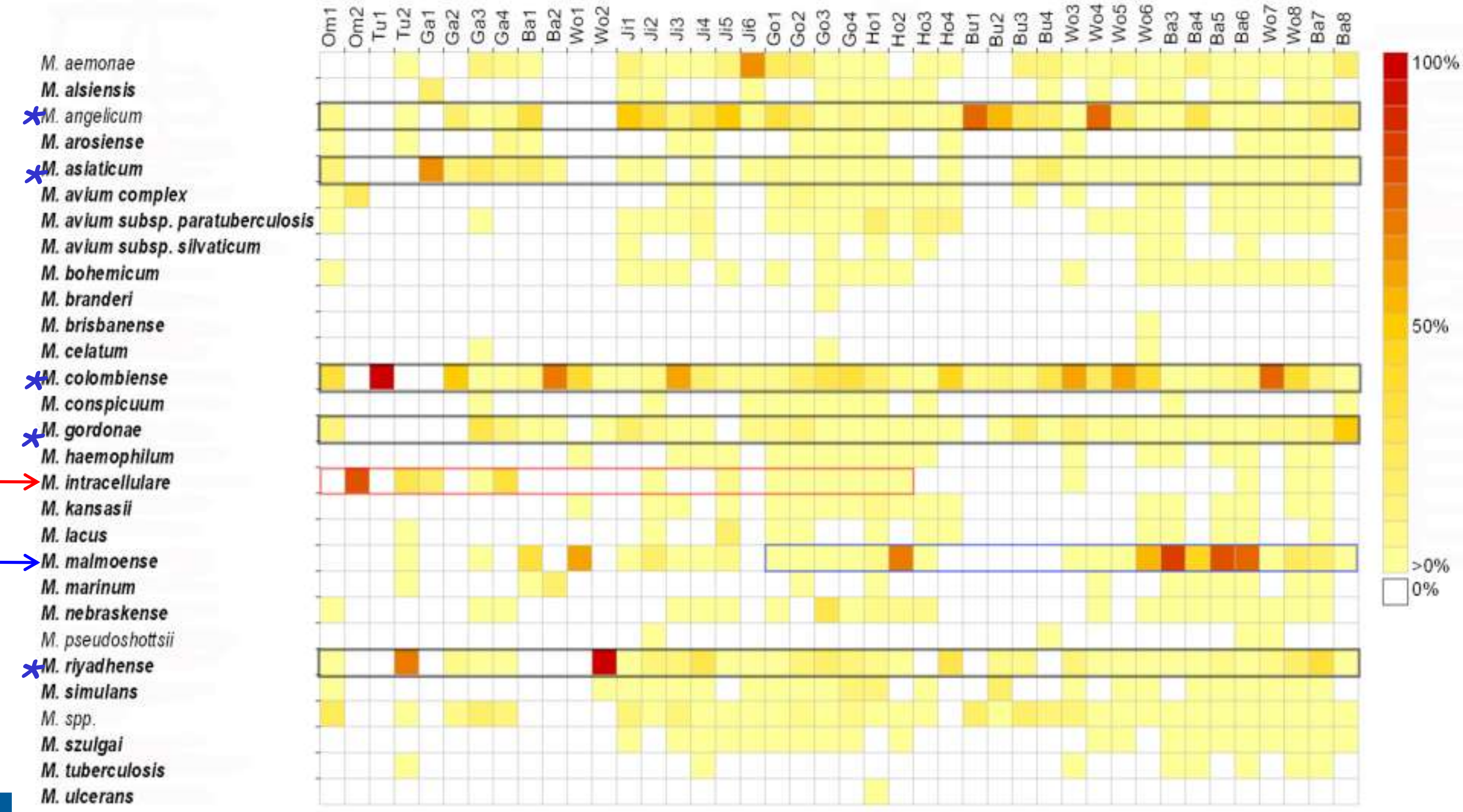
Slow-growing EM in Ethiopian soils

Alpha rarefaction



Diversity of slow-growing EM in soils

Hot arid lowlands → Warm agricultural midlands → Tepid highlands

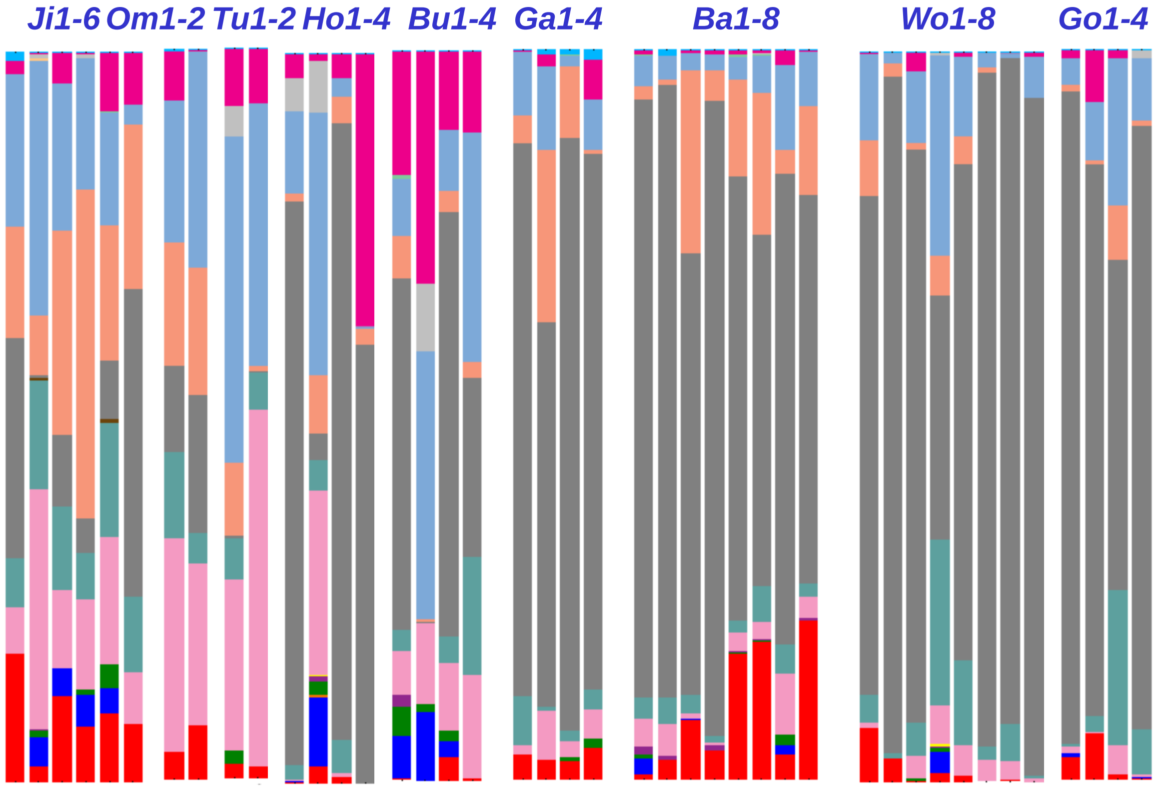


BLAST results at a 97% identity and 90% alignment coverage

Slow-growing EM water samples:

taxonomy plots

Bale
Butajira
Jinka
Gambella
Gondar
Hossana
Woldiya

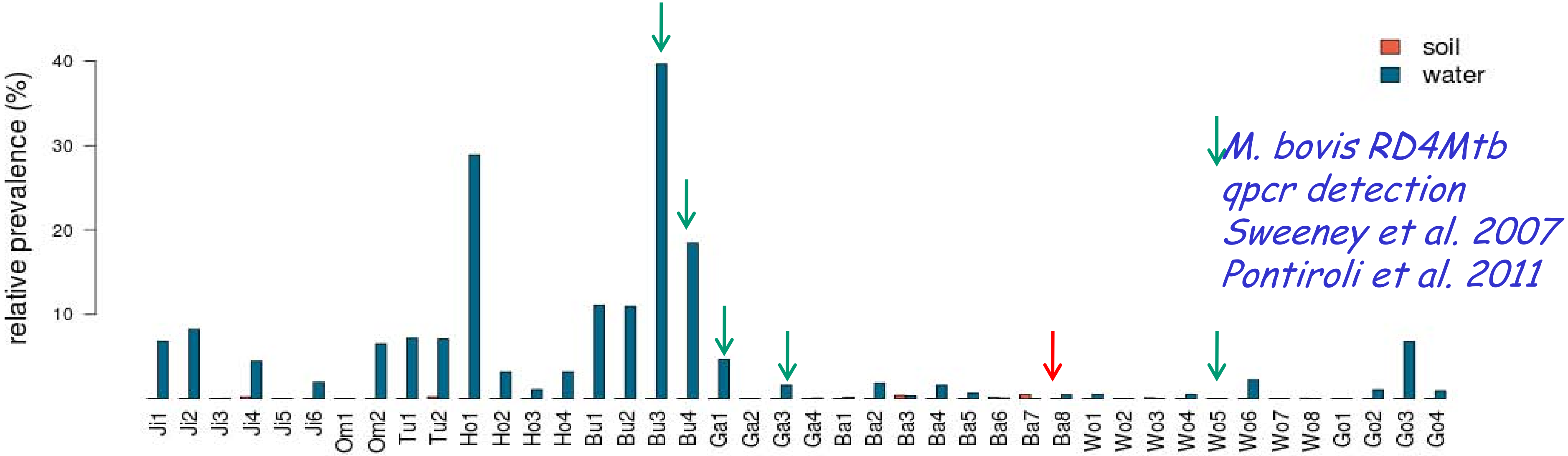


- 🟥 Mycobacterium aemonae
- 🟦 Mycobacterium alsiensis
- 🟧 Mycobacterium angelicum
- 🟩 Mycobacterium arosiense
- 🟪 Mycobacterium asiaticum
- 🟨 Mycobacterium bohemicum
- ■ Mycobacterium bovis
- ■ Mycobacterium colombiense
- ■ Mycobacterium conspicuum
- 🟫 Mycobacterium gastri
- ■ Mycobacterium gordonae
- 🟩 Mycobacterium intracellulare
- ■ Mycobacterium malmoense
- ■ Mycobacterium riyadhense
- ■ Mycobacterium shimoidei
- ■ Mycobacterium sp.20612A
- ■ Mycobacterium sp.C43a
- ■ Mycobacterium sp.HSC2034
- ■ Mycobacterium sp.MB69
- ■ Mycobacterium tuberculosis
- ■ Mycobacterium ulcerans

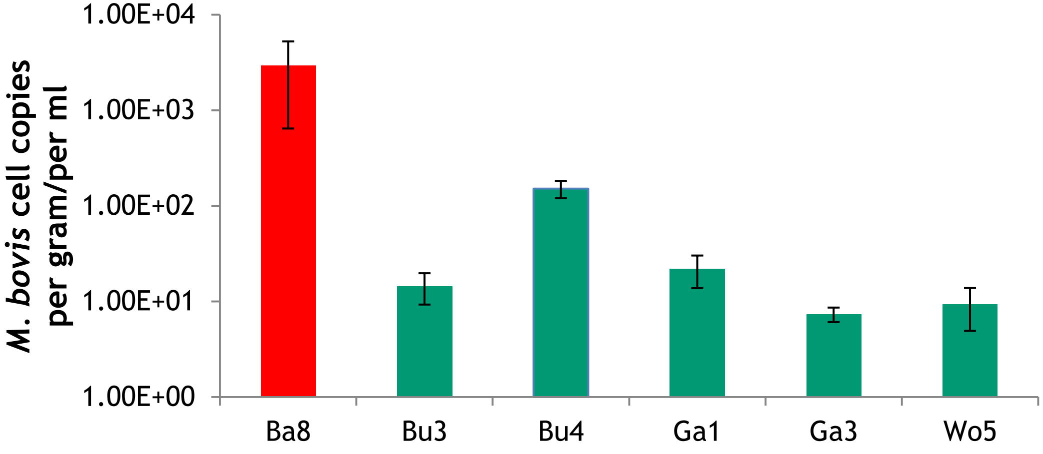
Higher prevalence of M. tuberculosis complex, and M. gordonae, M. malmoense, M. conspicuum , M. colombiense

Detection of *M. bovis* in Ethiopian samples

M. tuberculosis complex pyrosequencing reads



M. bovis RD4Mtb qpcr detection
Sweeney et al. 2007
Pontiroli et al. 2011



*Data suggests **water** sources are a potential reservoir of M. bovis infection*
Human-Environment-Livestock-interface

2.38% (1/42) of soils were positive for M. bovis
11.90% (5/42) of water samples were positive for M. bovis

ACKNOWLEDGEMENTS

Co-workers

Chitinases: Ashley Johnson-Rollings, Helena Wright

Antibiotics: Paris Laskaris, Nikos Kyratsous

Resistome: Will Gaze, Lihong Zhang, Greg Amos

Community analysis: Tanya Khera collaborator Orin Courtenay

Collaborators:

Peter Hawkey, University of Birmingham, UK

David Pearce BAS, Cambridge, UK

Brian Oakley, CDC Atlanta, USA

Girum Erenso, Armaue Hansen Research Institute, Ethiopia

Wyeth

bbsrc

NATURAL ENVIRONMENT RESEARCH COUNCIL

SEVENTH FRAMEWORK PROGRAMME

ACTAPHARM

Novo Nordisk