Comparative genomics of intestinal protozoa

Simone M. Cacciò

European Union Reference Laboratory for Parasites

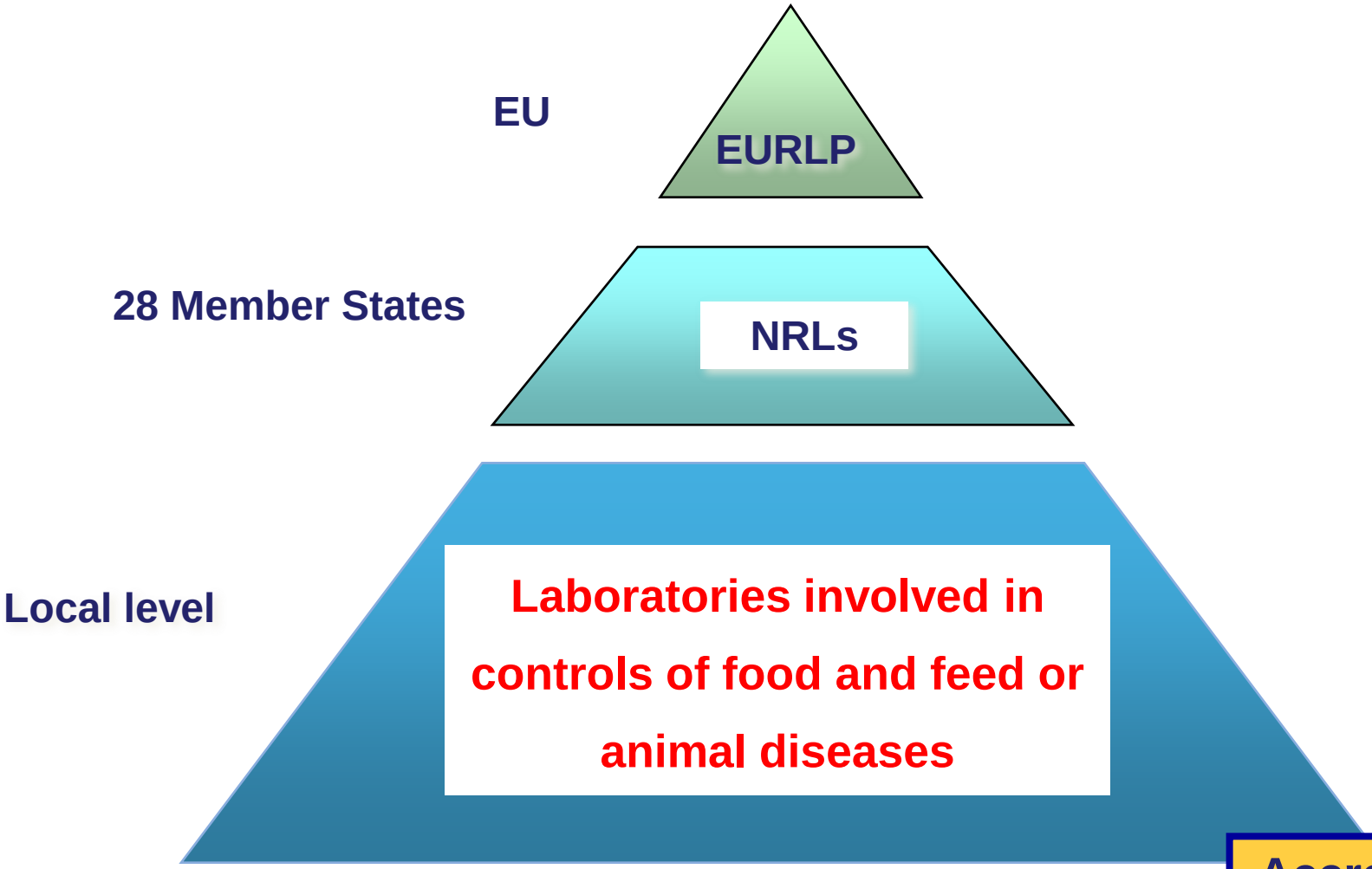
Istituto Superiore di Sanità, Rome, Italy

12th Annual Workshop of the National Reference Laboratories for *E. coli* in the EU, Rome, 12-13 October, 2017

The EURL for Parasites

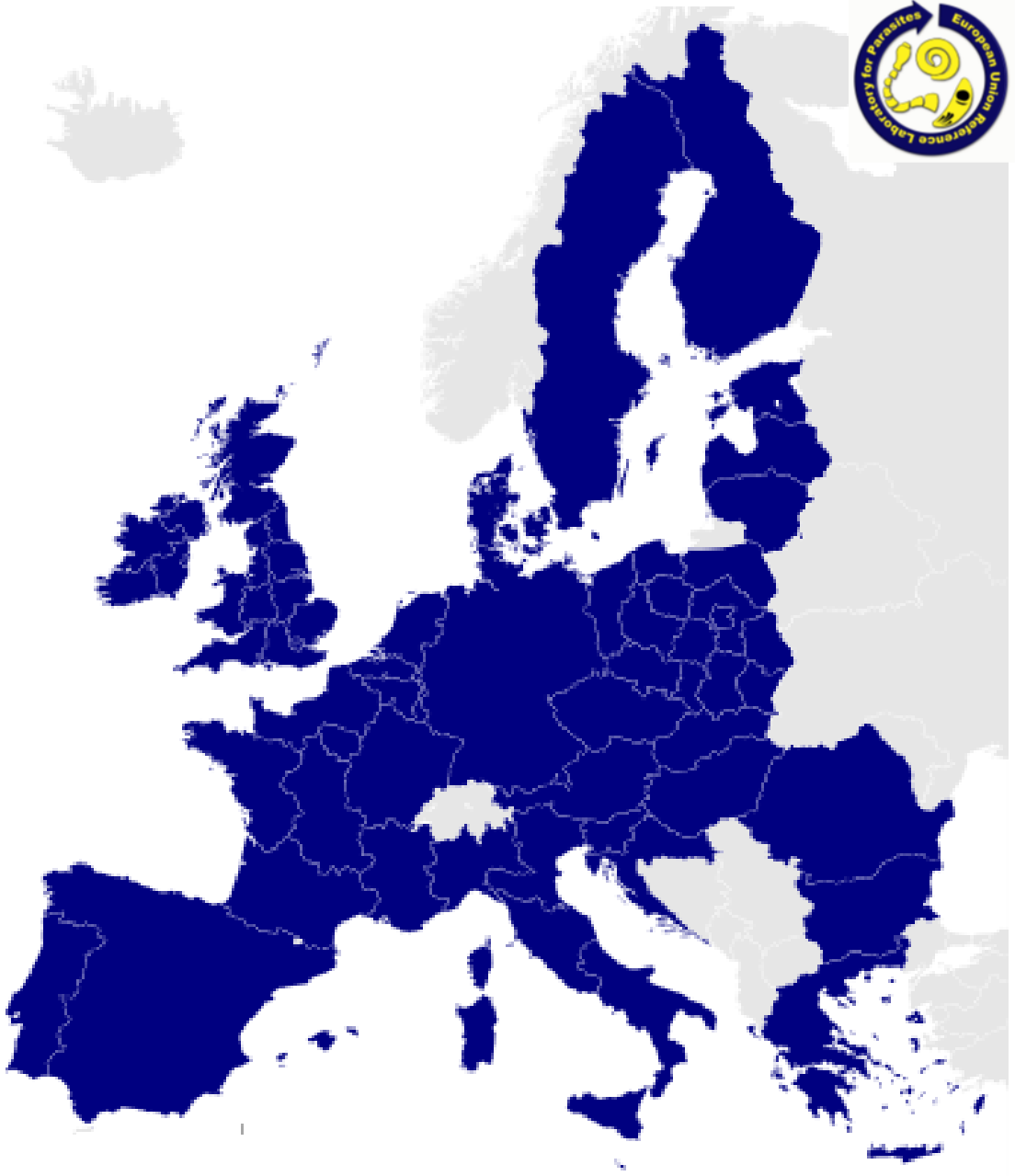
- The DG SANCO of the European Commission has appointed the EURLP at the Istituto Superiore di Sanità since 2006
- The EULP has been accredited by the Italian accreditation body ACCREDIA:
  - since July 2006 for ISO/IEC 17025:2005 which specifies the general requirements for the competence to carry out tests and/or calibrations
  - since March 2014 for ISO/IEC 17043, which specifies the general requirements for proficiency testing

The Network of Reference Laboratories for Parasites



Accreditation ISO 17025

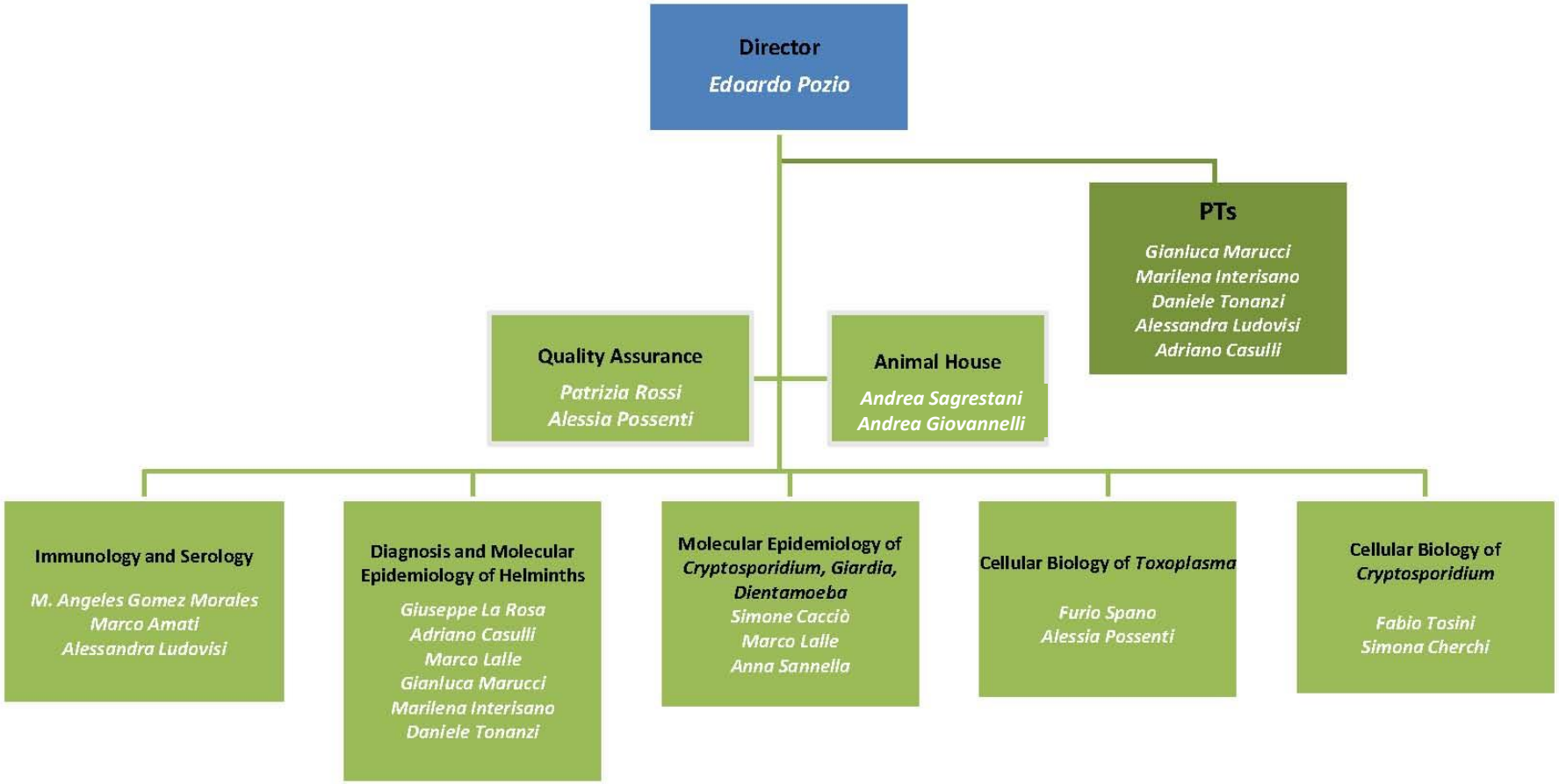
National Reference Laboratories for Parasites



Each of the 28 MS has appointed 1-4 NRL for parasites

European Union Reference Laboratory for Parasites

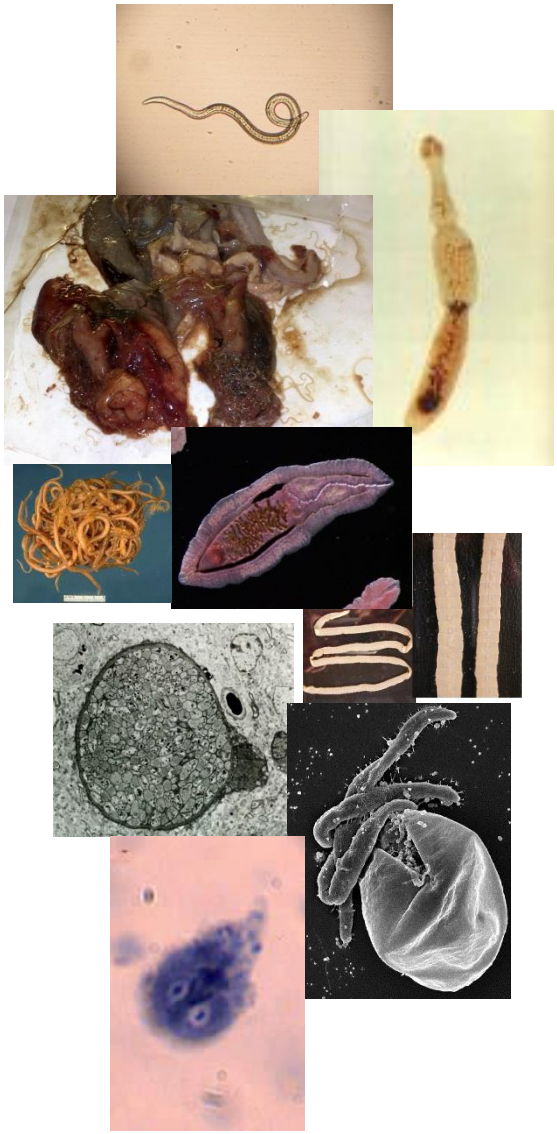
Department of Infectious, Parasitic and Immunomediated Diseases
Unit of Gastroenteric and Tissue Parasitic Diseases
Istituto Superiore di Sanità



The EURLP website: www.iss.it/crlp/index.php

The EURL for Parasites

- Target parasites:
  - **Helminths**
    - *Trichinella*
    - *Echinococcus*
    - *Anisakis*
    - *Pseudoterranova*
    - *Opisthorchis*
    - *Diphyllobotrium*
    - *Ascaris*
    - *Toxocara*
    - and other foodborne helminths
  - **Protozoa**
    - *Toxoplasma*
    - *Giardia*
    - *Cryptosporidium*
    - *Sarcosystis*
    - *Dientamoeba*
    - and other foodborne protozoa



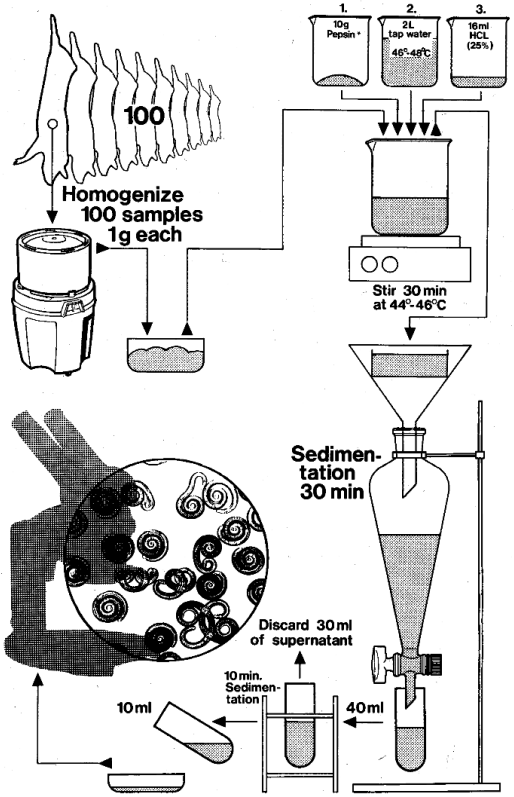
The EURLP activities- 1

- **Development of diagnostic methods**
  - Methods for the direct and indirect detection and identification of parasites of public health importance in the MS (*Trichinella*, *Echinococcus*, *Anisakis, Opisthorchis*, *Cryptosporidium*, *Giardia, Toxoplasma*) have been developed, validated and accredited

The EURLP activities- 2

Diagnostic methods for *Trichinella*

- Direct detection method
  - Detection of *Trichinella* larvae in muscle tissues
- indirect detection method
  - Detection of anti-*Trichinella* sp. antibodies in swine sera by indirect ELISA and Western blot
  - Detection of anti-*Trichinella* sp. antibodies in human sera by indirect ELISA and Western blot



The EURLP activities- 3

Development of diagnostic methods

Molecular tests for parasite identification

Identification of ***Trichinella*** sp. larvae at the species level by Multiplex PCR

Identification of **Anisakidae** larvae at the species level by PCR/RFLP

Identification of **Anisakidae** larvae at the species level by Multiplex PCR

Identification of ***Echinococcus granulosus*** complex hydatid cysts at species/genotype level by PCR and sequencing

Identification of ***Cryptosporidium*** oocysts at the species level by PCR/RFLP

Identification of ***Opisthorchis*** *spp.* eggs by PCR

Identification of ***Giardia duodenalis*** cysts at the Assemblage level by PCR/RFLP

Identification of ***Giardia duodenalis*** Assemblages A and B by PCR

Identification of ***Toxoplasma gondii*** DNA in feed and food by LAMP

The EURLP activities - 4

- **Production and collection of Reference Material**
  - Given that international standard materials for diagnosing foodborne parasitic zoonoses are currently not available, one of the main activity of the EURLP is to develop reference materials
    - Animal and human reference sera
    - Parasite antigens (crude and excretory/secretory antigens)
    - Parasite reference strains
    - DNA from reference strains

- **Diagnostic activities**
  - An average of 2000 diagnoses per year are performed at the EURLP, upon request from NRLs

- **Epidemiological investigations**
  - The EULP personnel support the personnel of NRLs on epidemiological investigations and diagnosis in the course of human and animal outbreaks due to foodborne parasites (*Trichinella*, *Opisthorchis*, *Anisakis*, *Cryptosporidium*, *Giardia*)

The EURLP activities- 5

- **Proficiency testing (PT) provider**

  NRLs of EU and of Candidate Countries undergo PT on foodborne parasites. After a PT is completed, a full evaluation report is prepared and provided to each laboratory. The results of the PT are discussed during the workshop. In 2018, the following PT will be organized:

  – PT on detection of *Trichinella* larvae in meat samples
  – PT on molecular identification of *Trichinella* larvae at the species level
  – PT on detection of *Echinococcus* adult worms in intestinal contents
  – PT on detection of Anisakidae larvae in fish fillets
  – PT on molecular identification of Anisakidae larvae at the species level
  – PT on molecular identification of Echinococcus at the species level
  – Ring trial on detection of anti-Toxoplasma IgG in serum samples of sheep, goats and pigs by commercial kits or in-house tests

- **Workshop**

  -The EURLP organizes a yearly workshop for the NRLs to discuss issues related to foodborne parasitic diseases in Europe, as well as the results of inter-laboratory comparison studies, the annual EURLP work program, and scientific news in the field of foodborne parasites. Every year an average of 70 persons attend the workshop.

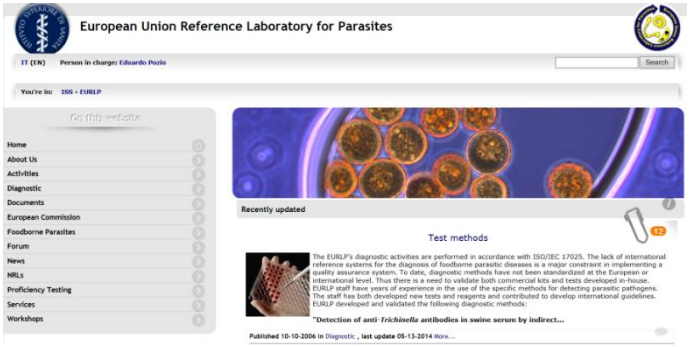
  -Participants are from the 28 EU countries and from associate countries (Former Yugoslav Republic of Macedonia, Iceland, Norway, Serbia, Switzerland)

The EURLP activities- 6

- **Organization of ring trials**
  - Ring trials are organized for evaluating diagnostic tests in European and extra-European laboratories
- **Training activities**
  - Personnel of NRLs and from developing countries are trained on diagnostic tests and epidemiological investigations for foodborne parasites
  - On average, 10-15 persons are trained per year

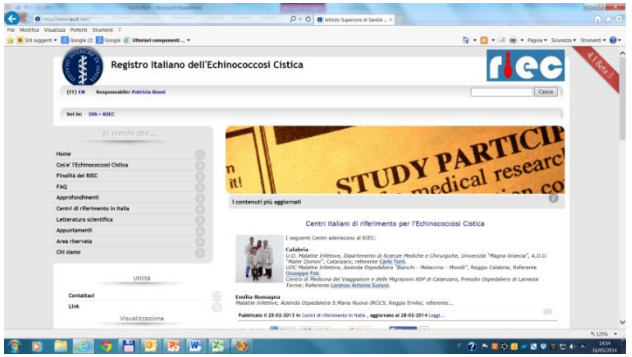
The EURLP activities- 7

- **Websites**
  - The EURLP website, with information on the activities of the Reference laboratory
  - (www.iss.it/crlp/)



  - International *Trichinella* Reference Center (ITRC), a repository of more than 5000 *Trichinella* isolates with information on species, host, locality of origin, year
  - www.iss.it/site/Trichinella/index.asp)

  - The Registry of Cystic Echinococcosis to collect clinical and epidemiological information on this zoonosis
  - (www.iss.it/riec)



The EURLP activities- 8

- **Research activities (2016-2017)**

- **ELISA to detect *anti-Trichinella* IgG in swine muscle juices**
- **Western blotting as a confirmatory test for ELISA-positive pig serum samples**
- **Coordinating the collection of *Trichinella* isolates in MS for microsatellite-based mapping**
- **Identification of taenid cestodae eggs in the definitive host (canids) faeces**
- **Identification of nematode larvae, different from those of the genus *Trichinella*, detected in muscles by artificial digestion**
- **Design and validation of informative typing schemes for *Cryptosporidium parvum* and *C. hominis***
- **Development of a molecular test for the identification of *Toxoplasma gondii* oocysts in fresh fruits and vegetables**

The EURLP activities- 9

Research activities (2018), based on NGS

Transcriptomics

- Current serological assays for *Trichinella* are based on the use of crude extracts or excretory/secretory antigens of muscular larvae. It is impossible to distinguish recent from old infections, in which only residual antibodies are present.
- The goal is to identify stage-specific antigens to be used for more sensitive and specific serologic assays.
- The strategy will be based on sequencing RNAs from adult worms, new borne larvae and muscular larvae
- Candidate antigens will be selected, expressed and tested with serum samples from experimentally and naturally infected pigs

The EURLP activities- 10

Research activities (2018), based on NGS

Genomics

- *Toxoplasma gondii* infects virtually all warm-blooded animals, including humans, livestock, marine mammals and birds. Of the three distinct clonal lineages, type II is by far the most prevalent in Europe.
- Genetic variation does occur in the type II population, but typing is still not fully informative.
- NGS will generate new genomic data for type II parasites.
- A mapping approach should be sufficient to reconstruct the new genomes and to identify highly polymorphic markers.
- A more robust and rationale genotyping scheme will be derived from NGS data.

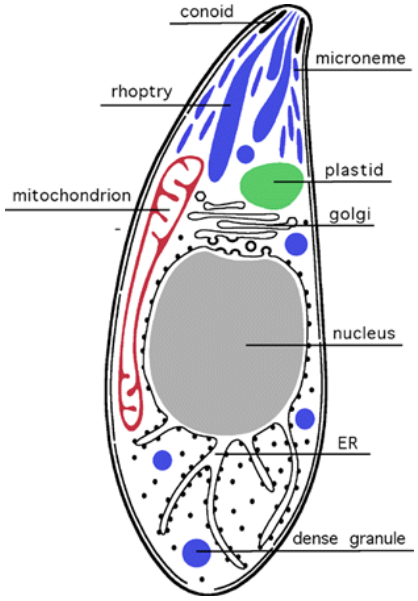
Current research based on NGS

NGS on protozoa (2015-2019)

- Comparative genomics of two protozoa, *Cryptosporidium* and *Giardia*.
- Metagenomics of intestinal protozoa

Cryptosporidium: taxonomy

- *Cryptosporidium* is included in phylum **Apicomplexa**, which comprises human and animal pathogens of great importance (*Plasmodium*, *Toxoplasma*)



All Apicomplexa are unicellular **parasites**, and most are **intracellular** or live in close contact with host cells.
They share the so-called **apical complex**, a sophisticated structure fundamental for host cell invasion, that gives the phylum its name.

Cryptosporidium: essential background

- Many species infect humans
- Globally distributed
- Complex epidemiology
- Massive waterborne outbreaks
- Immunocompromised at high risk
- Pediatric infection
- Disease burden in developing countries
- Few treatment options and no vaccine
- Lack of (simple) animal models

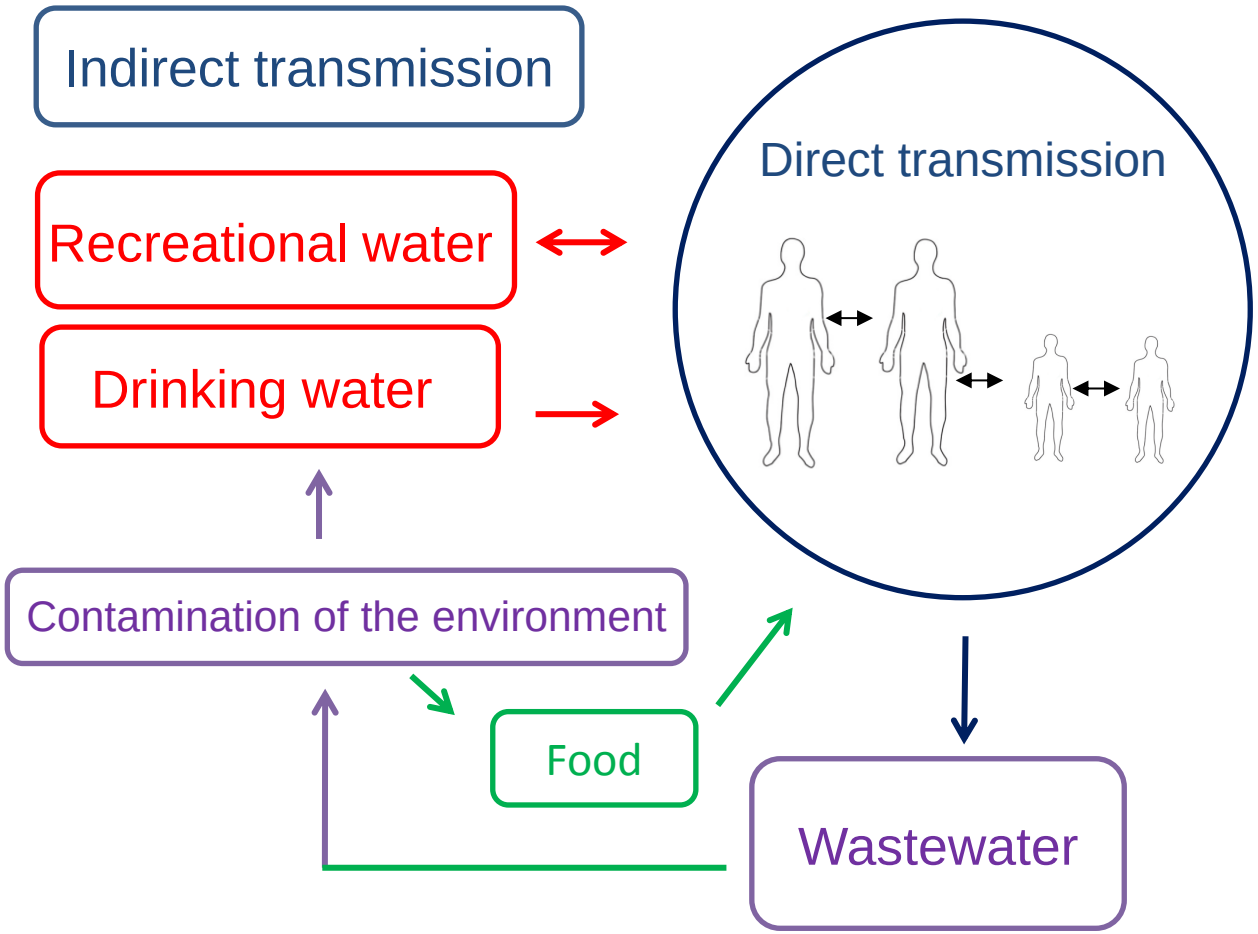
In young children, the parasitic infection cryptosporidiosis is one of four leading causes of severe diarrhoea.

Time to tackle cryptosporidiosis

The little-studied parasite *Cryptosporidium* is a major threat to infants.

Cryptosporidium hominis

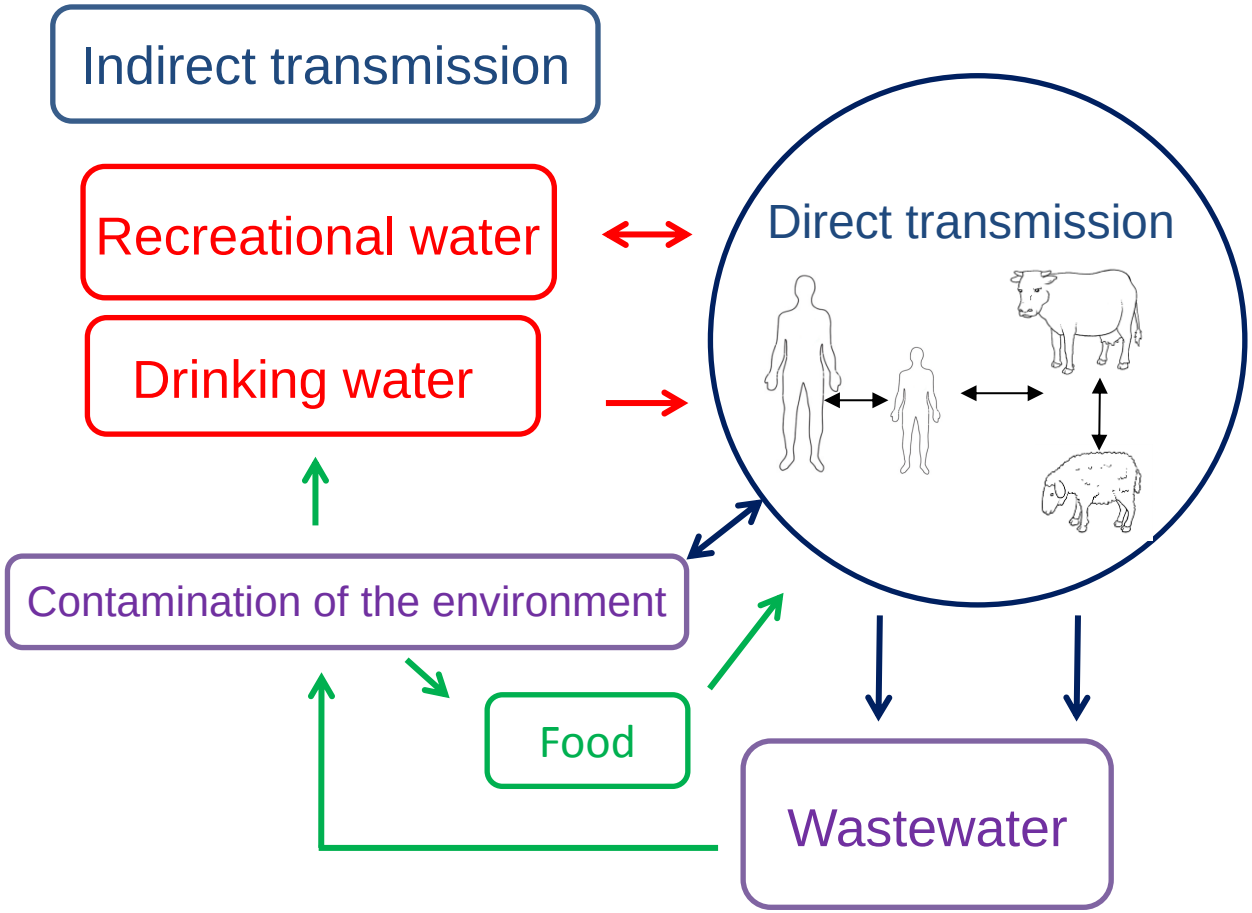
Major EU subtype:
IbA10G2



Main risk factors:

- Immune deficiencies
- Travel
- Drinking contaminated water
- Toiletting or changing young children
- Attendance at childcare centers
- Contact with another person with diarrhoea

Cryptosporidium parvum



Major EU subtypes:

IIaA15G2R1

IIaA15G1R1

IIaA17G1R1

IlaA19G1R1

IIaA19G2R1

IIaA18G3R1

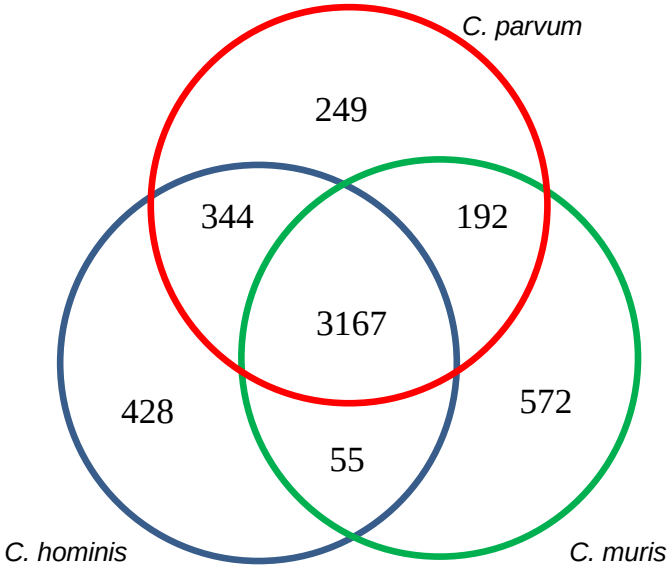
IId subtypes

Main risk factors:

- Contact with animals or dung, for example at petting farms
- Drinking contaminated water
- Immune deficiencies

Genomics: hard facts

- Small genome (9 Mb)
- Extremely streamlined metabolic pathways
- Organized in 8 linear chromosomes
- 75% annotated as protein-coding (40% hypothetical proteins)
- *C. hominis* and *C. parvum* genomes are largely syntenic



Genome comparisons: what questions?

- How much variation exists at the genome level?
- How similar are genomes from outbreak isolates?
- It is possible to infer transmission routes?
- How much recombination?
- Virulence factors, genes under selection
- Search for highly polymorphic regions to improve classical genotyping

A challenging work

- *Cryptosporidium* can not be grown *in vitro*. So, what we have is what is in the sample
- In the sample, target organisms are largely outnumbered by non-target organisms, thus purification is necessary
- A single oocyst contain about 40 femtograms (or 10^{-15}g) of genomic DNA, thus large numbers of purified oocysts are needed, and this is not always feasible

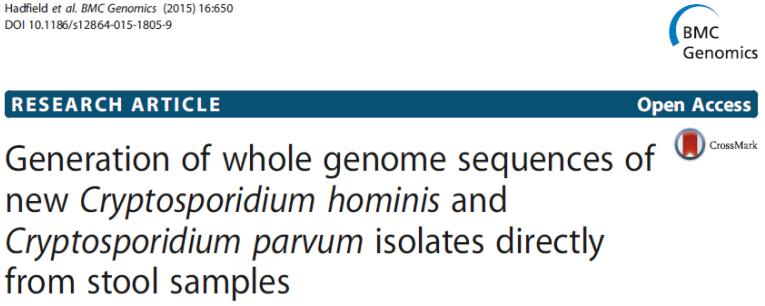

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Note

It's a dirty job — A robust method for the purification and *de novo* genome assembly of *Cryptosporidium* from clinical material


Hadfield *et al. BMC Genomics* (2015) 16:650
DOI 10.1186/s12864-015-1805-9

BMC Genomics

RESEARCH ARTICLE Open Access

Generation of whole genome sequences of new *Cryptosporidium hominis* and *Cryptosporidium parvum* isolates directly from stool samples

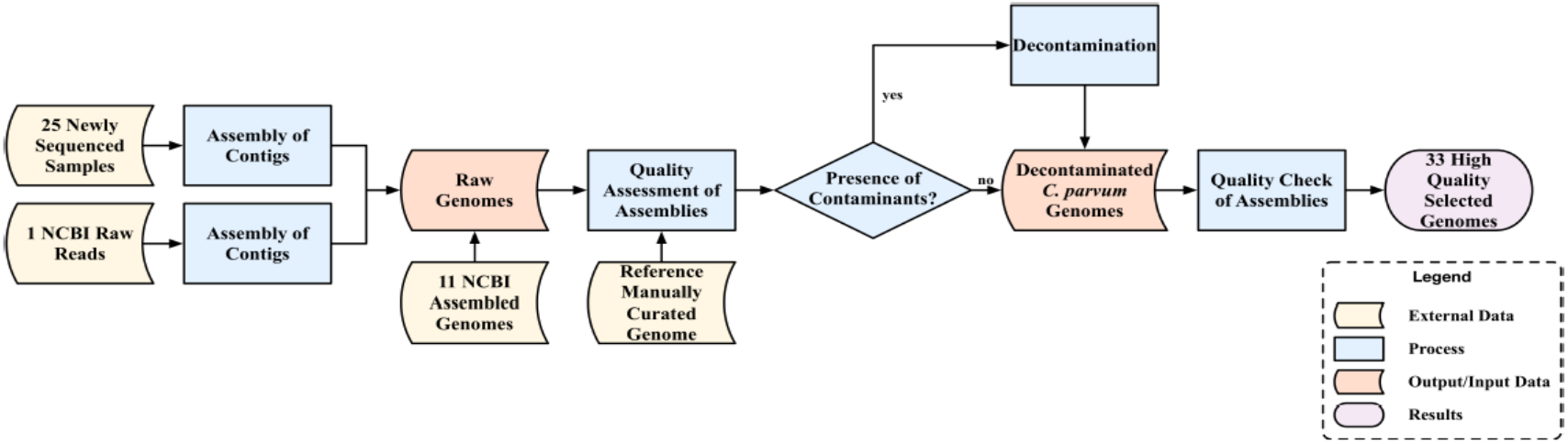
Where are we?

- We are actively collaborating with other groups to collect samples of both human and animal origin
- We have optimized the process of oocyst purification (immuno-magnetic separation) and DNA extraction (para-magnetic beads)
- We have sequenced more than 40 genomes, and twice as many are being processed.
- We are developing a pipeline for data processing and analysis, in collaboration with the European Bioinformatics Institute and the Centre for Integrative Biology (University of Trento)
- The analysis of 33 *C. parvum* genomes (14 from humans, 8 from calves, 5 from goat kids and 6 from lambs) is under way

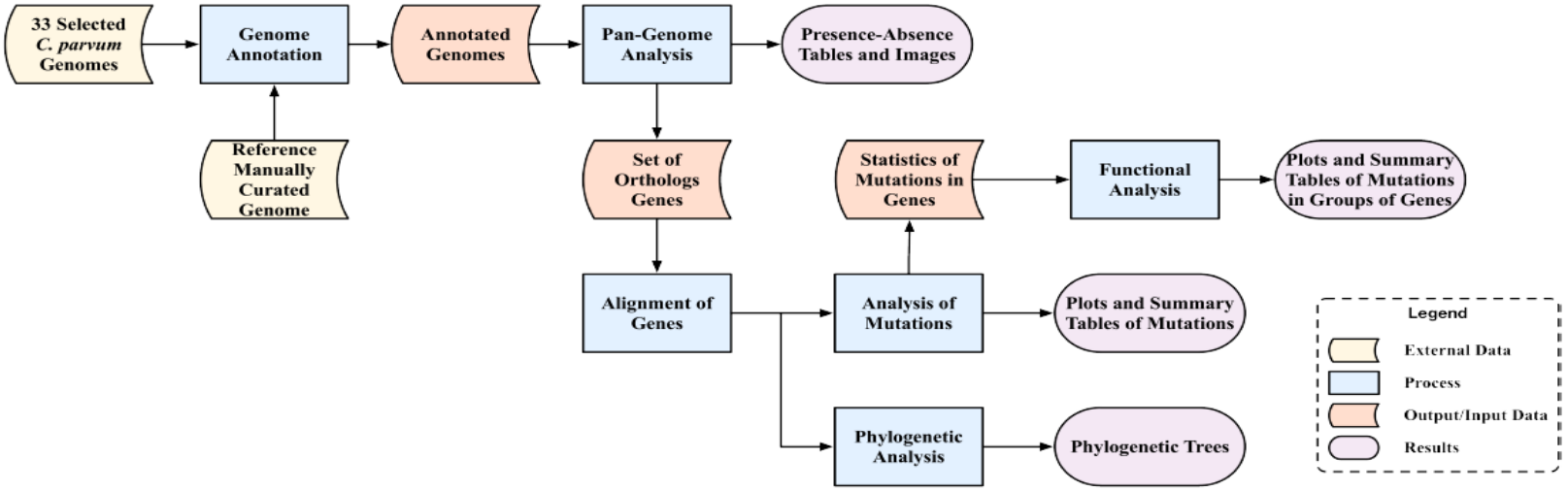
Preliminary analysis of *C. parvum* genomes

- We are actively collaborating with other groups to collect samples of both human and animal origin
- We have optimized the process of oocyst purification (immuno-magnetic separation) and DNA extraction (para-magnetic beads)
- We have sequenced more than 30 genomes, and twice as many are being processed.
- We are developing a pipeline for data processing and analysis, in collaboration with the European Bioinformatics Institute and the Centre for Integrative Biology (University of Trento)

Assembly, decontamination and quality assessement

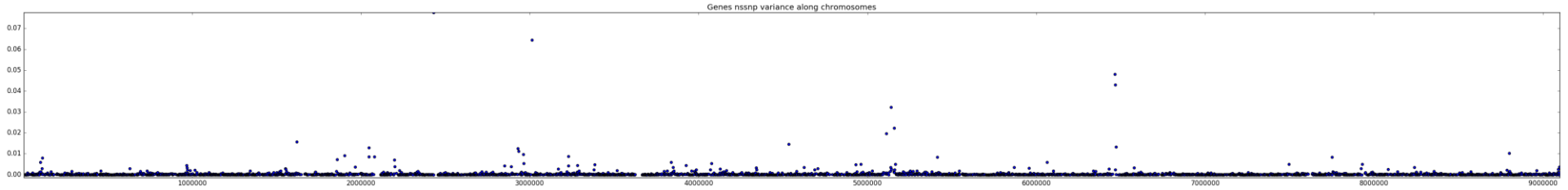


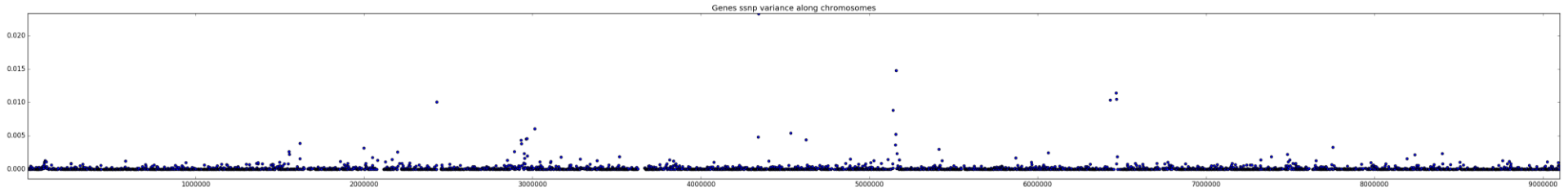
Genome annotation and analysis



Preliminary analysis of *C. parvum* genomes

Distribution of non-synonymous SNPs across the chromosomes



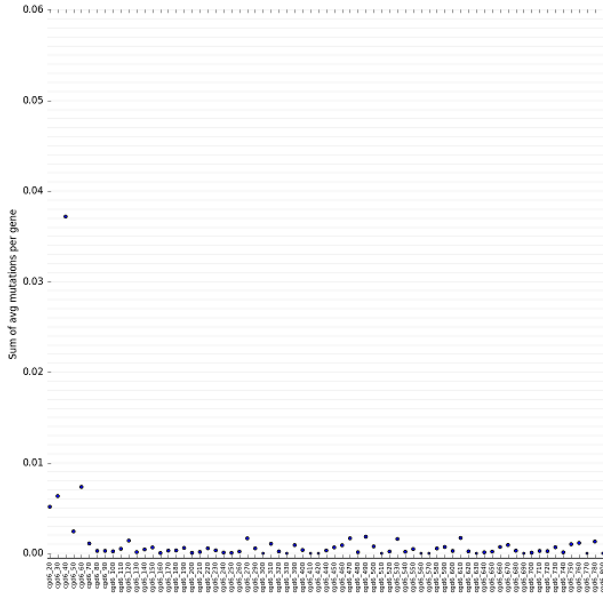


Distribution of synonymous SNPs across the chromosomes

Key message: **Similar distribution of Non-Syn and Syn mutations**
Variable genes often located in sub-telomeric regions

This confirms and extends result already published

Which genes are there?



Chromosome 6

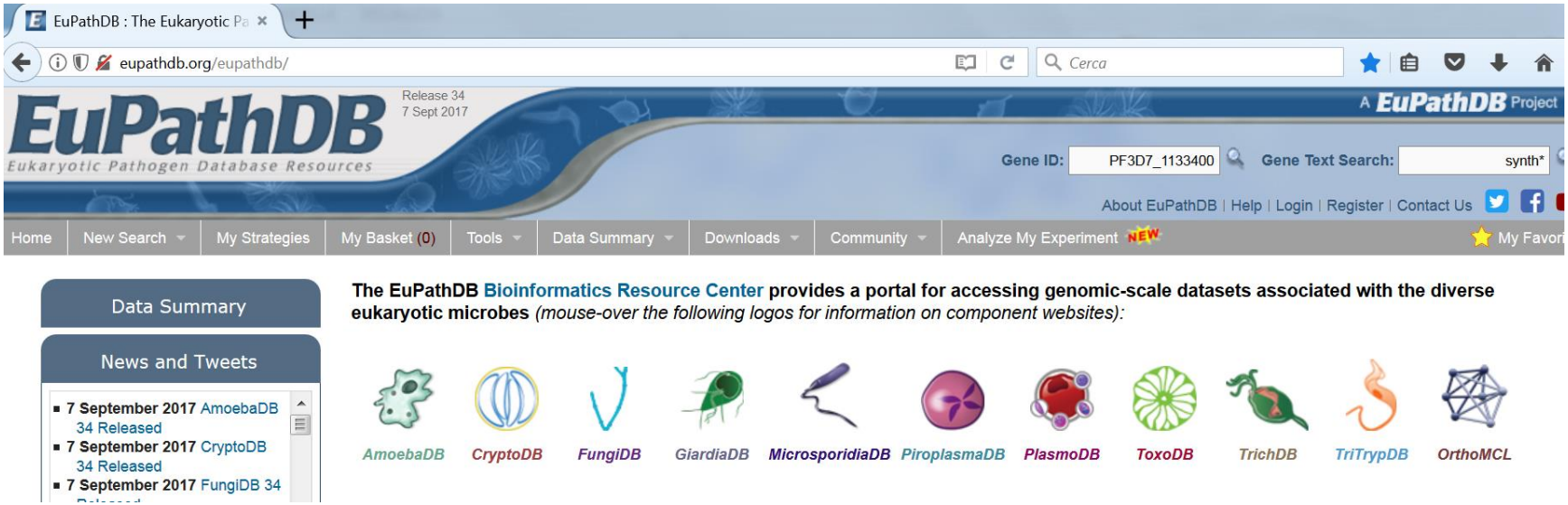
- conserved hypothetical protein
- hypothetical protein with signal peptide and proline stretch at C-terminus

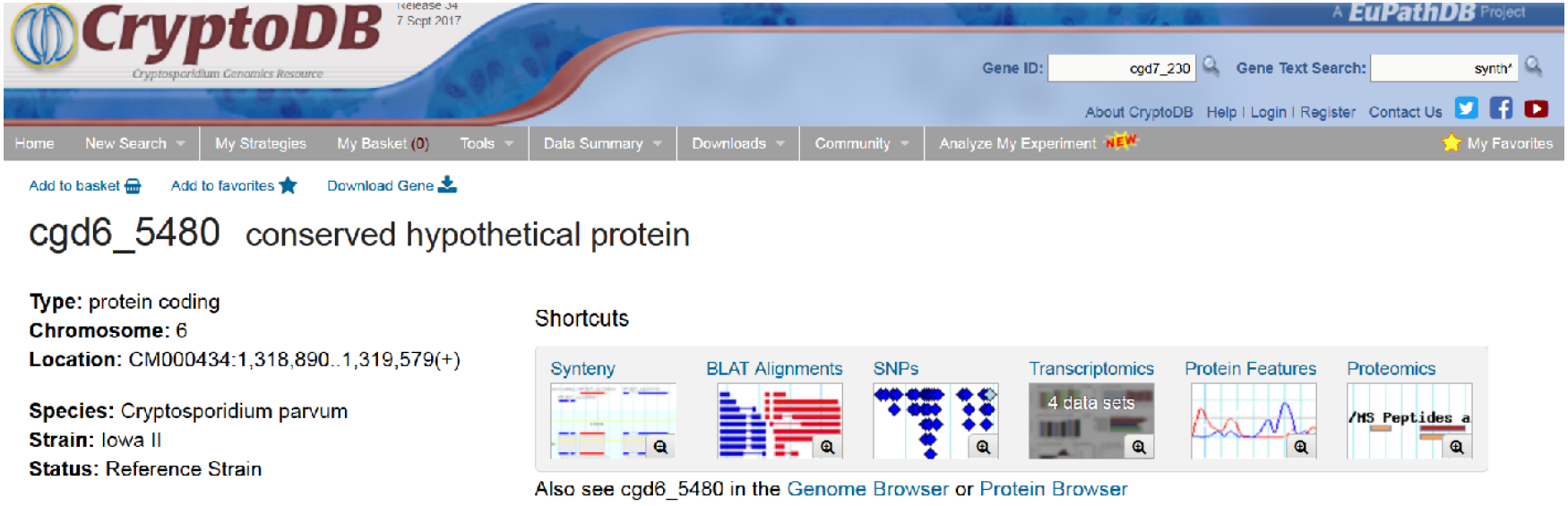
- conserved hypothetical protein
- conserved hypothetical protein
- **telomeric insulinase-like protease with signal peptide**

More generally, in the 10% most variable genes, many **hypothetical proteins**, **mucins** and **proteins with signal peptides** are found

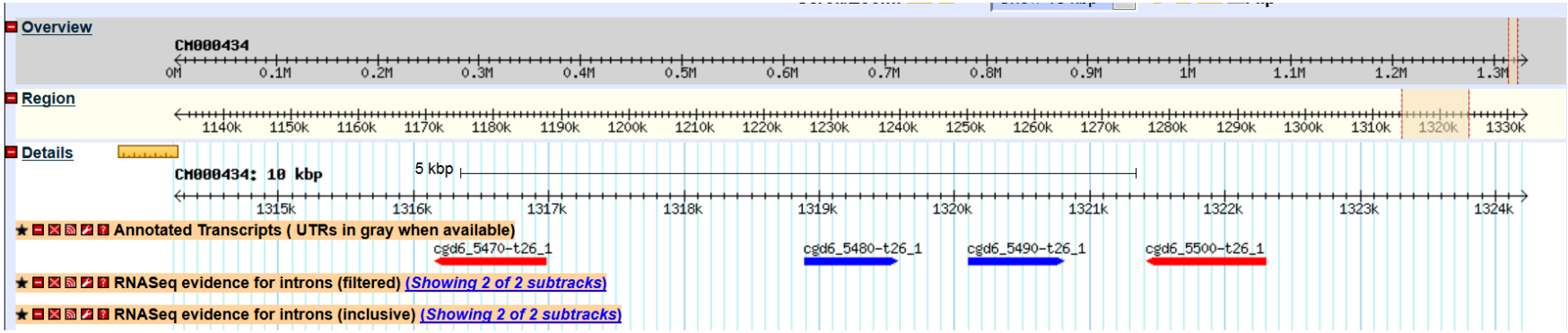
Cgd6_5480: telomeric insulinase-like protease with signal peptide

A closer look: EuPathDB





A closer look: EuPathDB



7 Orthology and synteny

Ortholog Group OG5_171489

▼ Orthologs and Paralogs within CryptoDB Data sets

Search this table... Showing 6 rows

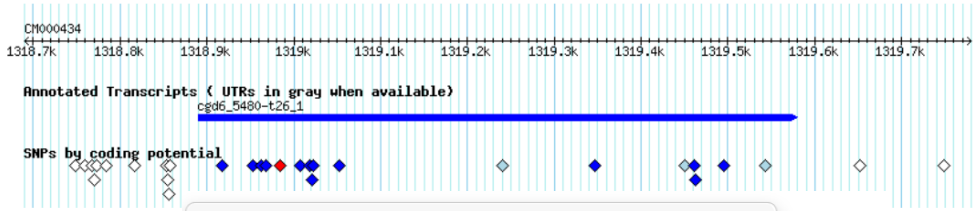
Gene	Organism	Product	is syntenic	has comments
CHUDEA6_5490	Cryptosporidium hominis UdeA01	unspecified product	yes	no
ChTU502y2012_317g0015	Cryptosporidium hominis isolate TU502_2012	hypothetical protein	no	no
Chro.50507	Cryptosporidium hominis TU502	hypothetical protein	no	no
cgd5_4600	Cryptosporidium parvum Iowa II	conserved hypothetical protein	no	no
cgd5_4610	Cryptosporidium parvum Iowa II	conserved hypothetical protein	no	no
cgd6_5490	Cryptosporidium parvum Iowa II	conserved hypothetical protein	no	no

8.2 DNA polymorphism

Non-Coding SNPs All Strains 0
NonSyn/Syn SNP Ratio All Strains 4.33
NonSynonymous SNPs All Strains 13

SNPs

View in genome browser

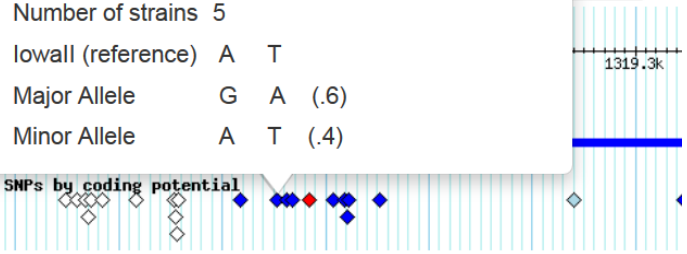


SNPs with Stop Codons All Strains 1
Synonymous SNPs All Strains 3
Total SNPs All Strains 17

Track details

SNP	NGS_SNP.CM000434.1318953
Location	1318953
Gene	cgd6_5480
Position in CDS	64
Position in protein	22
Type	Coding (non-synonymous)
Number of strains	5
Iowall (reference)	A T
Major Allele	G A (.6)
Minor Allele	A T (.4)

owser

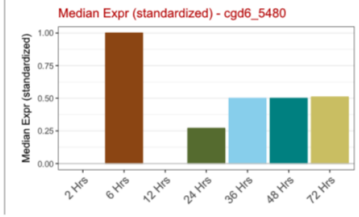


9 Transcriptomics

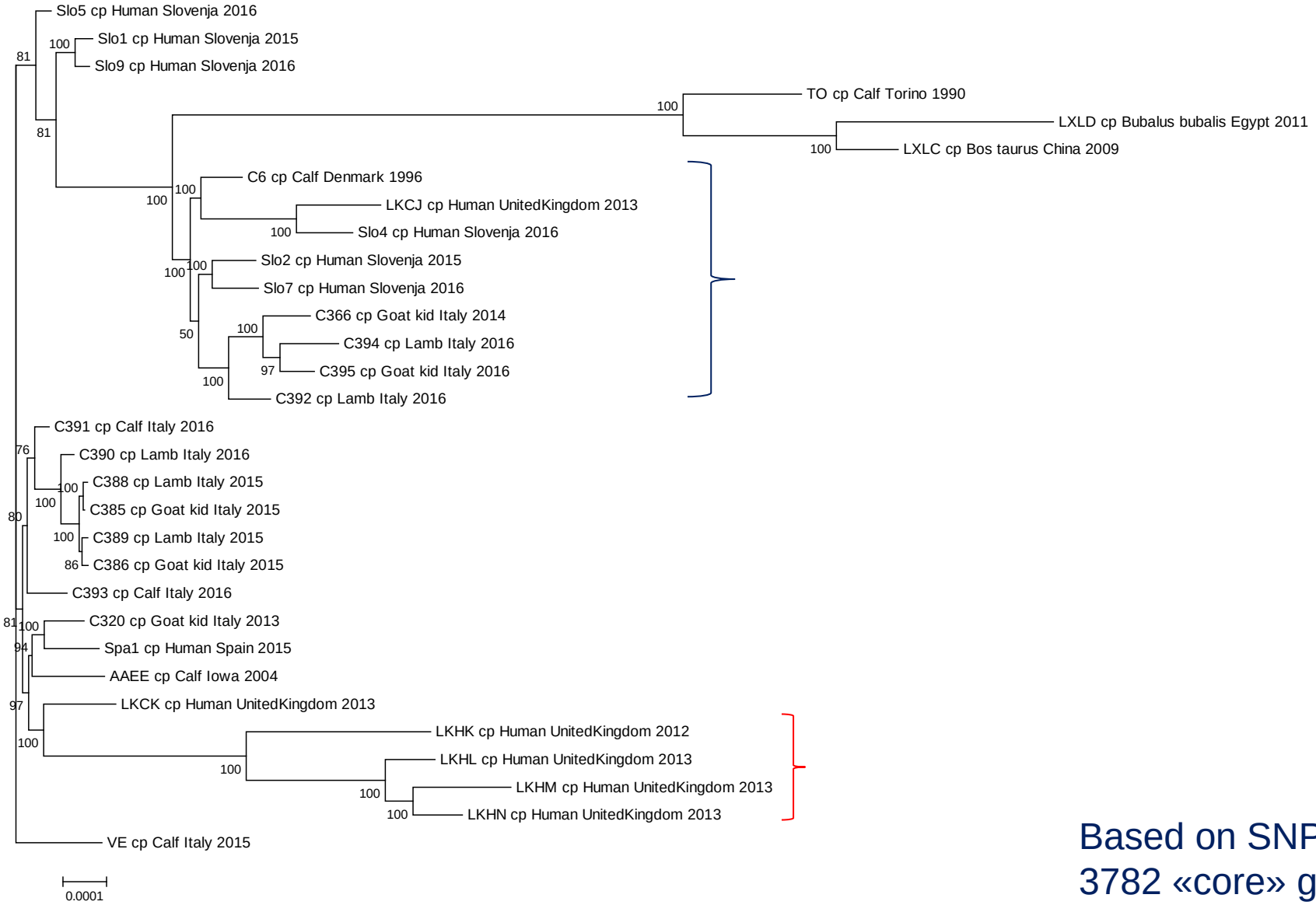
Transcript Expression Data sets

Search this table... Showing 4 rows

Preview	Name	Summary	Attribution	Assay Type
	Expression profiling of life cycle stages post-infection	Expression profiling of *Cryptosporidium parvum* during *in vitro* development determined by high throughput RT-PCR.		RT-PCR



Phylogenetic analysis



Based on SNPs in 3782 «core» genes

Metagenomics case study: *Blastocystis*

Why *Blastocystis*?

1) availability of reference genome sequences
2) high prevalence in humans, worldwide
3) pathogenicity controversial
4) Nine variants (subtypes) associated with human infection

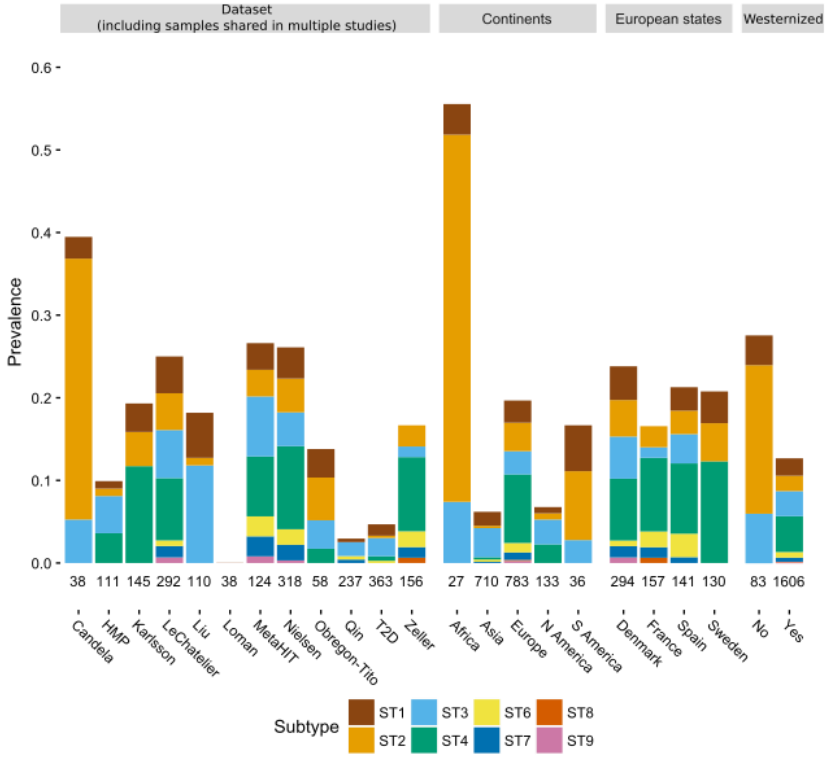
In collaboration with Nicola Segata, Francesco Beghini, Edoardo Pasolli (Centre for Integrative Biology, University of Trento, Italy)
Lorenza Putignani (Pediatric Hospital Bambino Gesù, Rome, Italy)

Outline of the study

- A total of 2,154 faecal microbiome samples from 1,689 subjects were used. The original 12 datasets included subjects from different geographic origins and/or different disease conditions.
- Metagenomic reads were mapped to reference genomes using the Bowtie2 aligner. The “breadth of coverage” is the number that represents the fraction of the reference genome covered by the metagenomic reads. Likewise, the “depth of coverage” is the average number of reads that cover each position of the genome.
- A detection threshold of >.10 (at least 10% of the genome represented) was defined.

Results: epidemiology

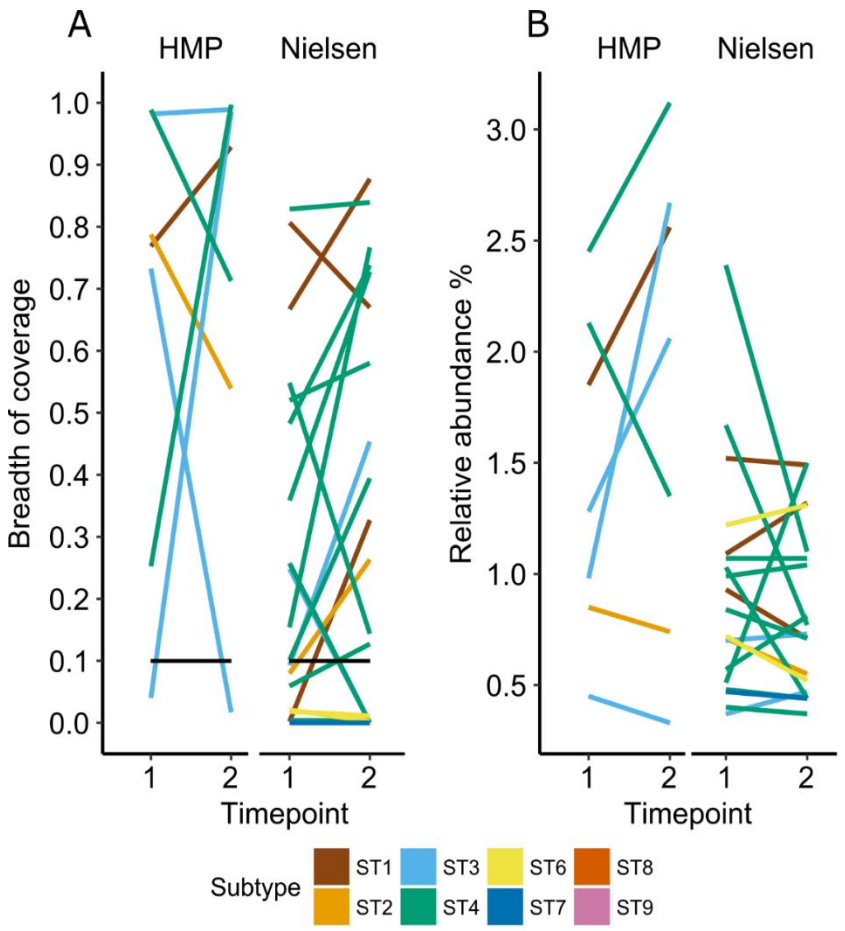
The parasite was detected in 321 samples (prevalence 15%) from subjects in China, Denmark, France, Mongolia, Norway, Peru, Spain, Sweden, Tanzania and USA. Evident non-random distribution of STs



Underrepresentation of ST4 outside Europe

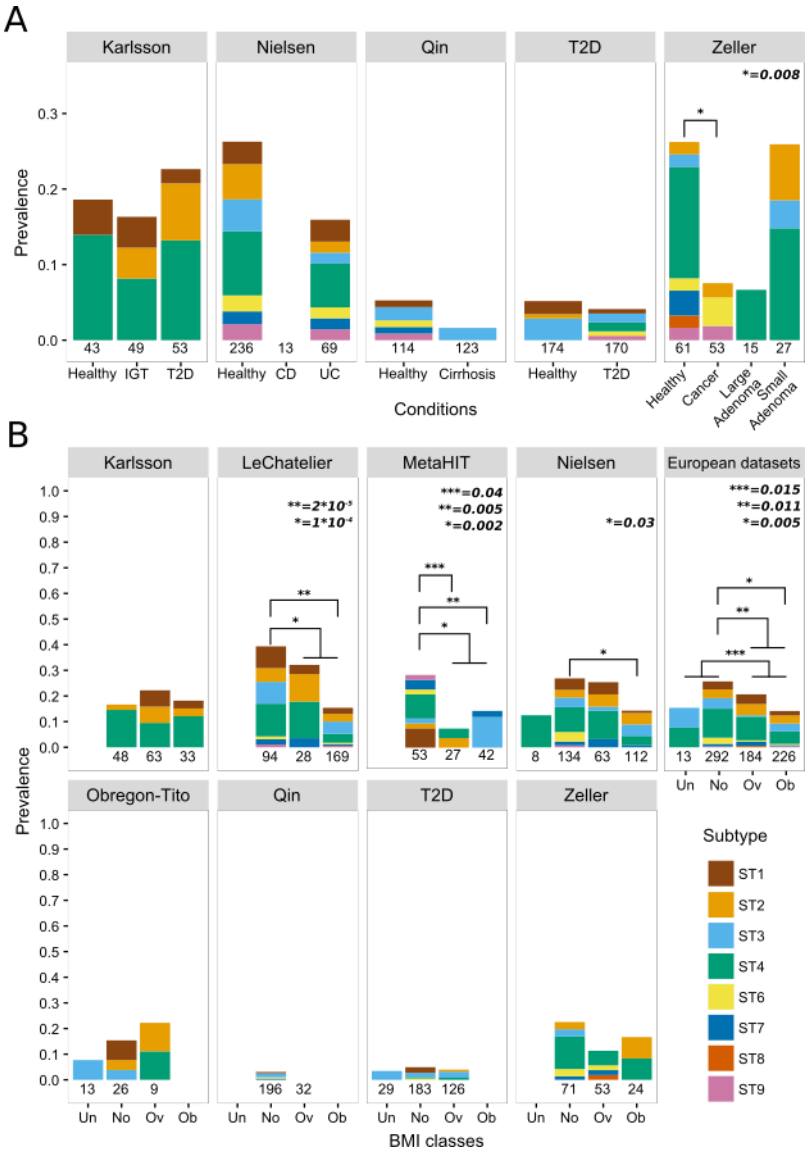
High prevalence of ST2 in non-westernized populations

Results: persistent colonization



- The parasite is able to persist for months, and that all the *Blastocystis* STs commonly associated with humans are able to stably colonize the gut
- The same ST is found at the two timepoints

Results: correlation with disease



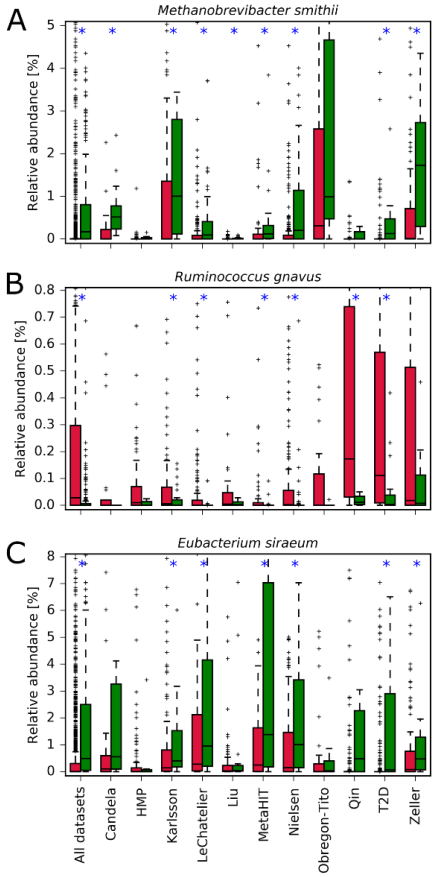
Higher prevalence in control groups than in diseased individuals

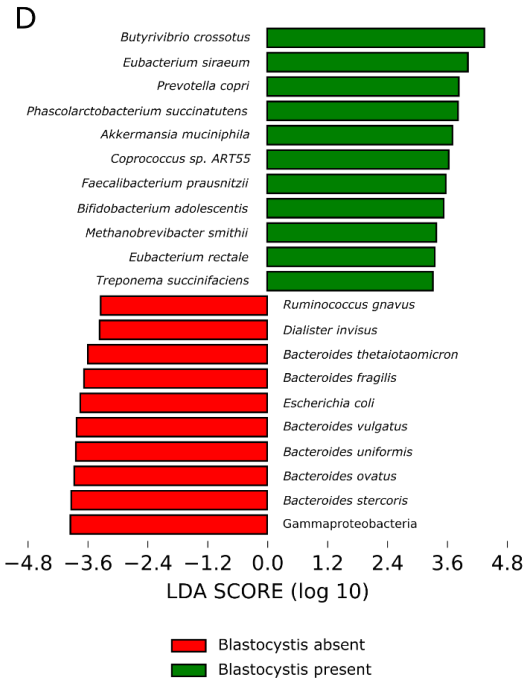
Inverse association between body mass index and *Blastocystis*

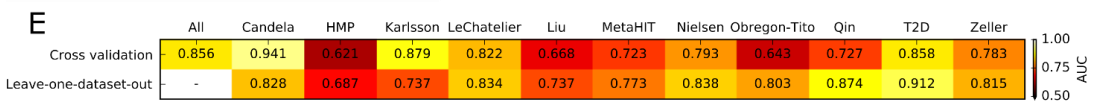
Is *Blastocystis* a member of a healthy gut microbiome and not a pathogen?

Results: interactions with gut microbiota

Strong co-occurrence with archaeal organisms (*Methanobrevibacter smithii*) and several bacterial species

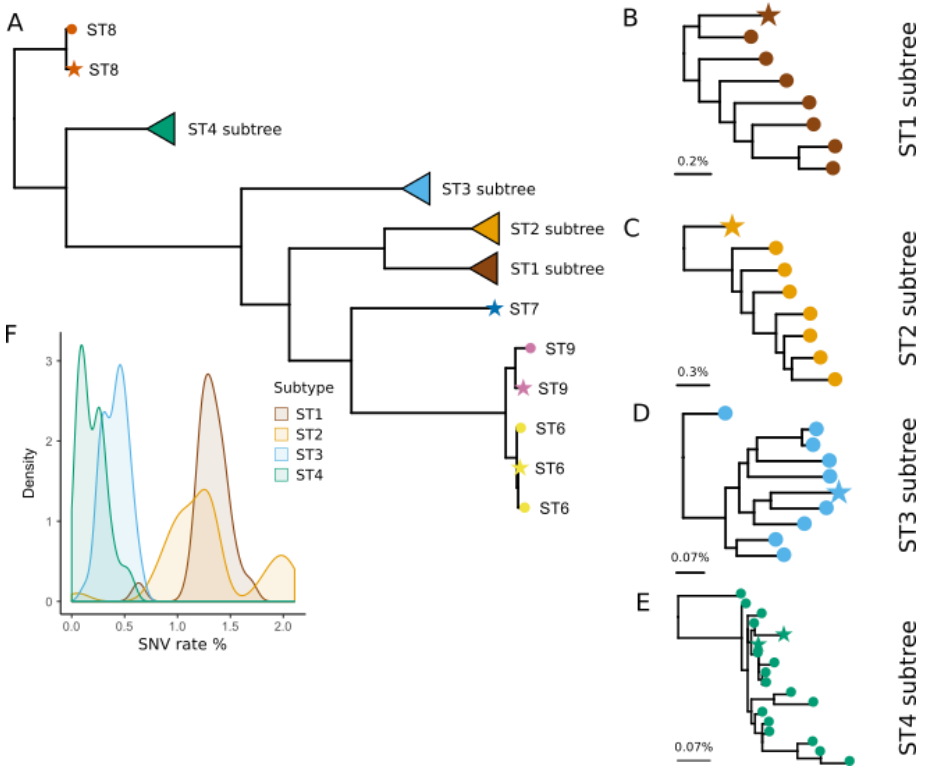


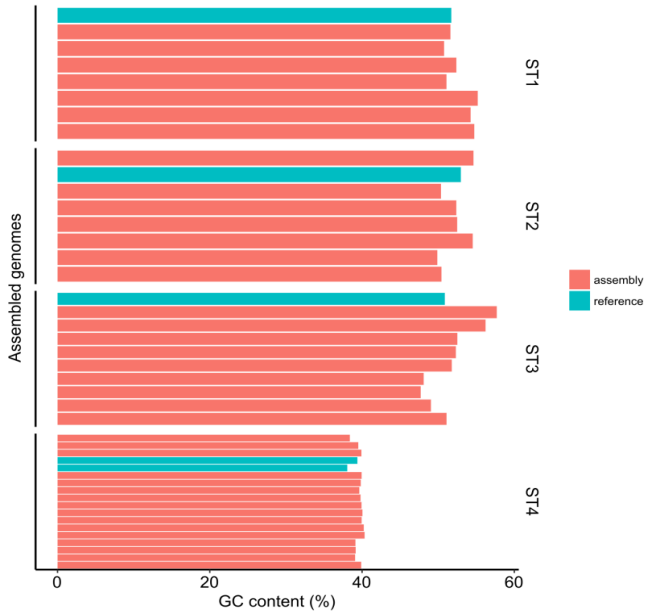




Results: new genomes from metagenomics data

It is possible to reconstruct high-quality *Blastocystis* assemblies from metagenomes, and explore, e.g, phylogenetic relationships and intra-genome diversity





In short

- We demonstrated the use of a pipeline to detect *Blastocystis* sequences in human gut metagenomics data
- We provide new insights into epidemiology, genomics, correlation with disease and with members of the microbiota
- We presented results supporting the concept that *Blastocystis* is more prevalent in a healthy gut rather than during dysbiosis


OPEN

The ISME Journal (2017), 1–16
www.nature.com/ismej

ORIGINAL ARTICLE

Large-scale comparative metagenomics of *Blastocystis*, a common member of the human gut microbiome

In conclusions

- Applications of NGS to the field of parasitology are still limited
- Specific technical challenges
- Not all isolates can be sequenced (what do we miss?)
- Few good reference genomes

- Parasites are very interesting biological systems
- Host-parasite interactions (metagenomics)