APPLICATION OF MODERN SEQUENCING TECHNOLOGIES TO INVESTIGATION OF OUTBREAKS CAUSED BY PROTOZOAN PARASITES IN ITALY

Sessione17

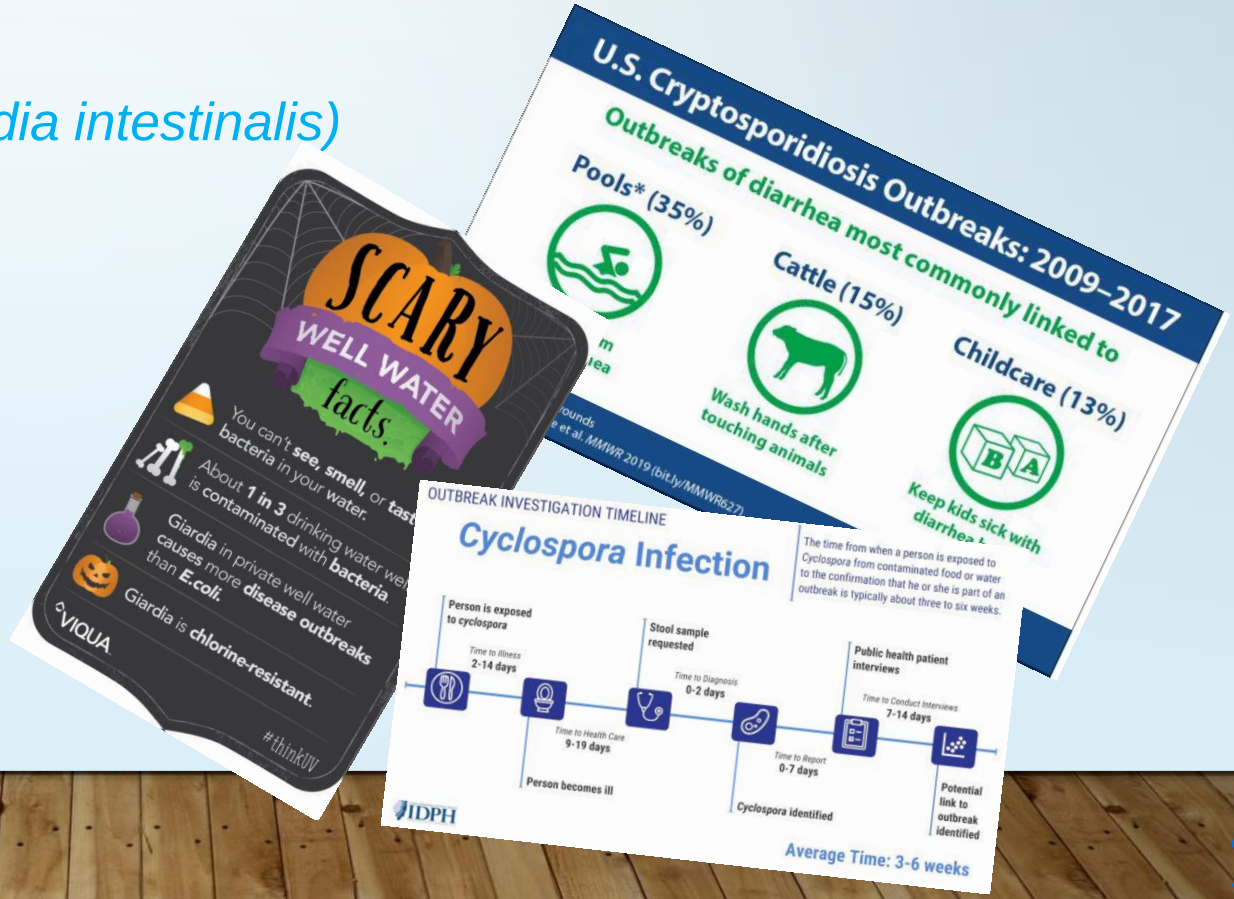
Marco Lalle

Istituto Superiore di Sanità, Roma, Italy

Most protozoa that infect the human enteric tract are characterized by having an environmentally stable stage such as a cyst or oocyst.

Cysts and oocysts confer protection from environmental factors, allowing these parasites to infect other susceptible hosts through either the **water or food-borne routes causing OUTBREAKS!**

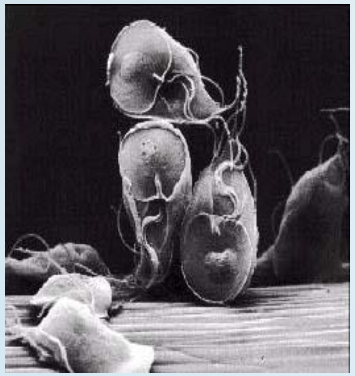
- *Cryptosporidium spp.*
- *Cyclospora cayetanensis*
- *Giardia duodenalis (syn. Giardia lamblia, Giardia intestinalis)*
- *Cystoisospora belli* (previously *Isospora belli)*
- *Toxoplasma gondii*
- *Entamoeba spp.*

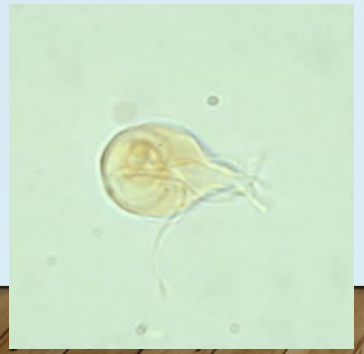


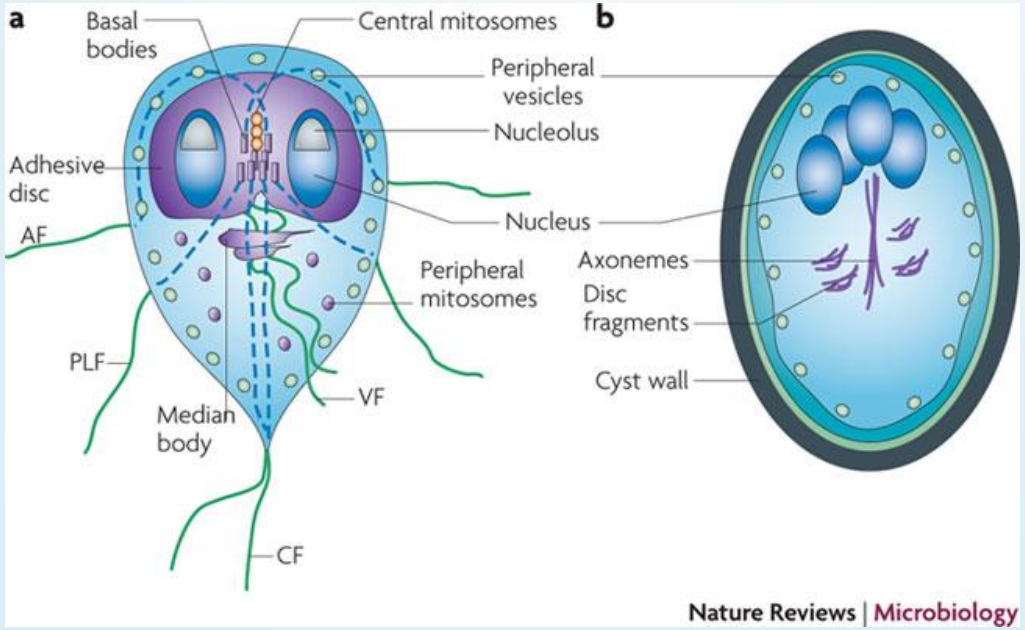
Giardia: one of these creatures

Giardia is a flagellated, unicellular eukaryote, that has a simple life cycle which comprises two stages:

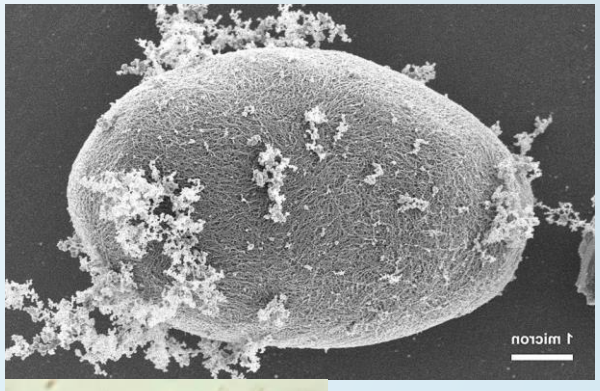
Trophozoite

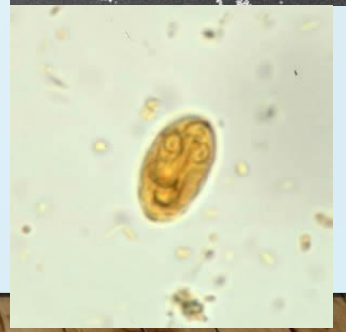




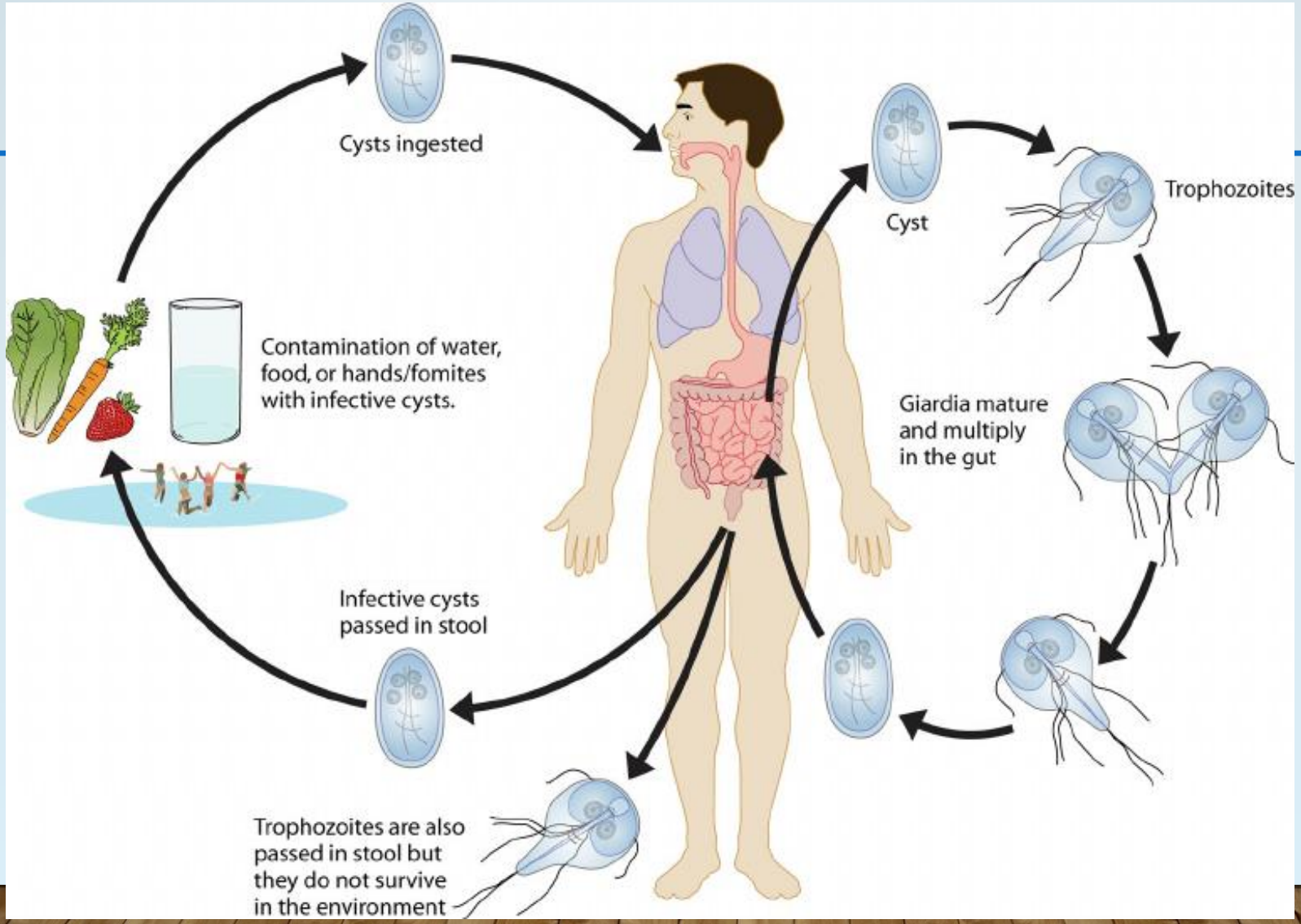


Cyst





Giardia life cycle



Source: Esch and Petersen. Clin Microbiol Rev. 2013 Jan;26(1):58-85

Taxonomy of *Giardia*

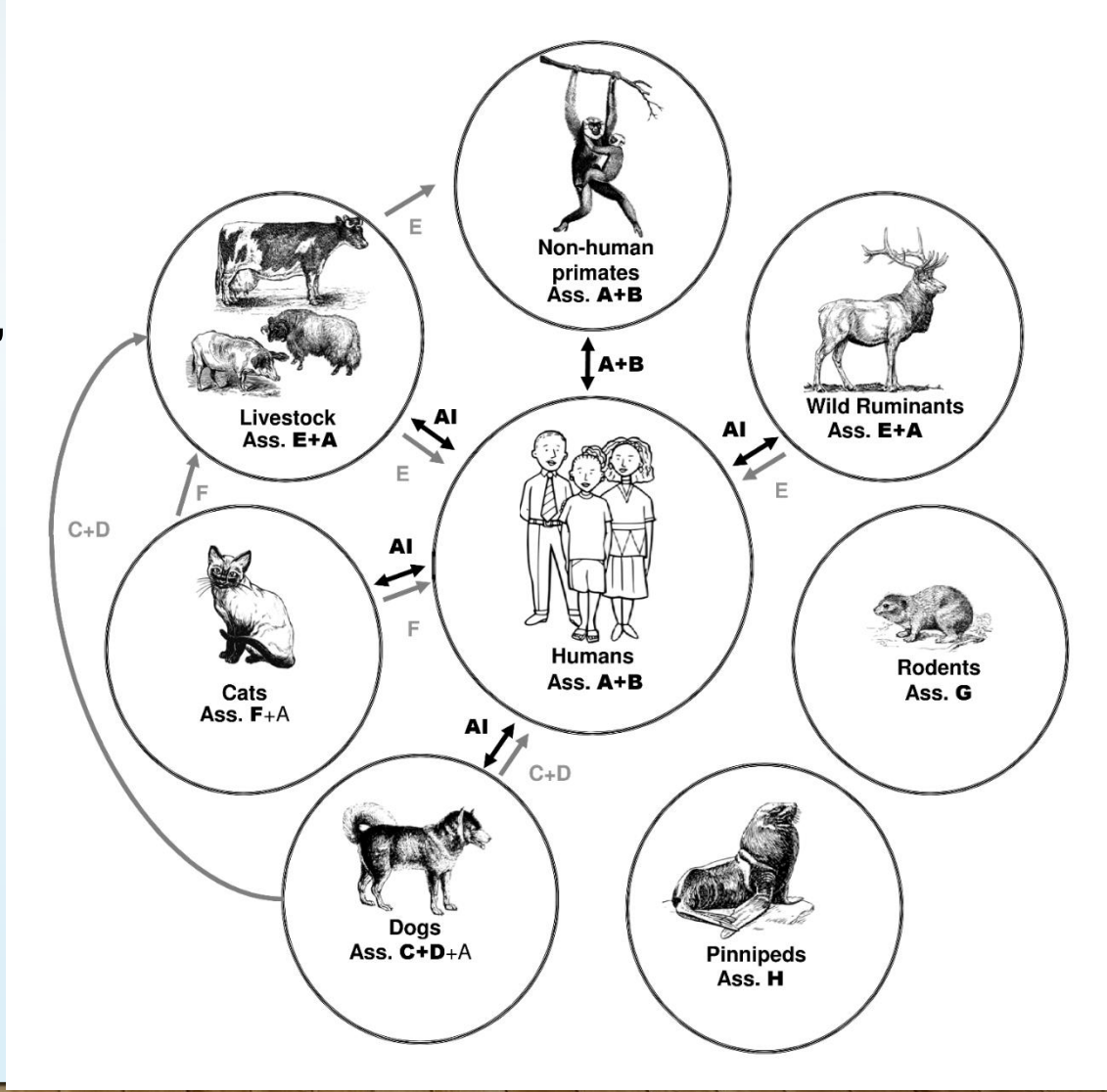
Species	Host
• *G. agilis*	Amphibians
• *G. ardeae*	Birds
• *G. psittaci*	Birds
• *G. muris*	Rodents
• *G. microti*	Rodents
• G. *cricetidarum*	Hamsters
• *G. peramelis*	Australian bandicoots
• ***G. duodenalis***	**Mammals (including man)**

However, *G. duodenalis* is not a single species.

If we look at the morphology of cysts from different hosts, there is no variability, but at DNA level a large amount of genetic variation is observed

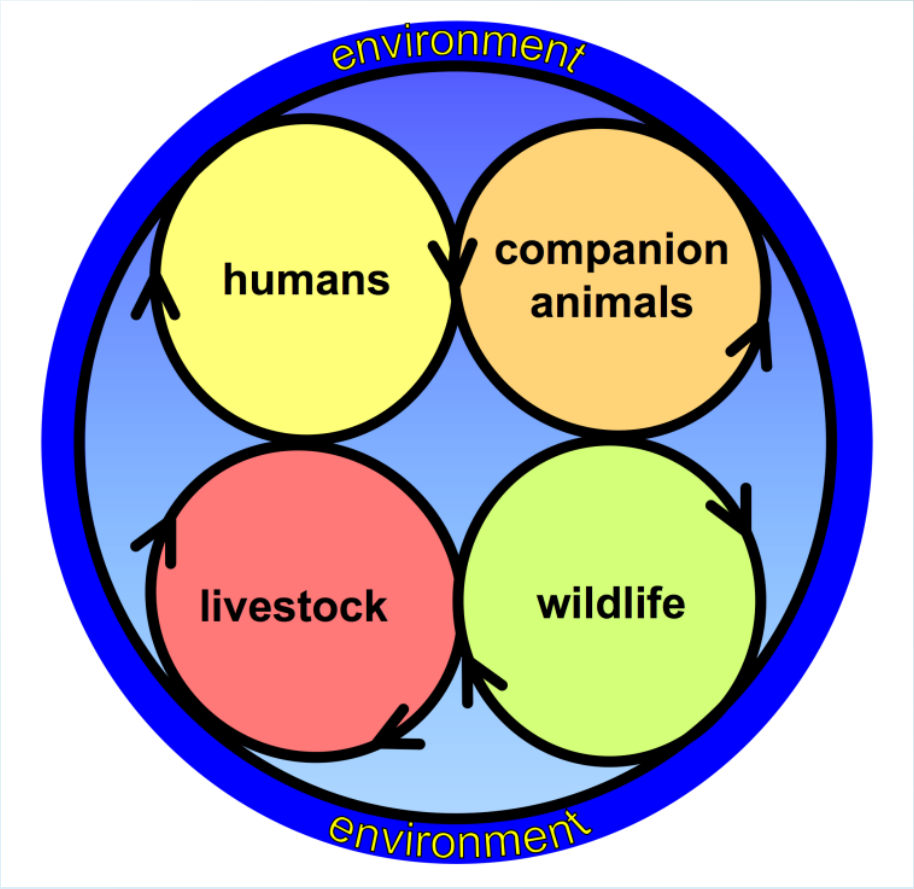
Giardia duodenalis

Genetic groups	**Host**
Assemblage A	➢**Man**, primates, livestock, pets, wildlife
Assemblage B	➢**Man**, primates, livestock, dog, beaver, horse, rat, muskrat
Assemblages C, D	➢dog, canids
Assemblage E	➢livestock
Assemblage F	➢cat
Assemblage G	➢rat, hamster
Assemblage H	➢marine mammals



General considerations on the epidemiology of human giardiasis

- Most cases are **waterborne** ➡ **outbreaks**
- Person-to-person transmission
- Zoonotic transmission (?)
- Foodborne transmission



The main biological features favouring transmission

- Ubiquitous, monoxenous, and able to cross-infect multiple species
- Large animal and human reservoirs
- Large numbers of infective cysts shed by infected individuals (i.e $2.5x10^7$cysts per inhabitant/year estimated in Netherland)
- Cysts immediately infectious upon excretion
- Small cysts (8–12 μm in length), which allows them to penetrate and survive water filters such as sand filters
- Cyst environmentally resistant:

❑can remain viable outside their hosts in aqueous environments for >3 months

❑at 4°C, remaining infective for 11 wk in water, 7 wk in soil, and 1 wk in cattle feces. In contrast, at −4 and 25°C cysts become non-infective after 1 and 2 wk, respectively

❑Resistance to chlorination

❑Biofilms may significantly contribute to the persistence of cysts in the environment, specifically in water distribution systems, as they act as transient or long-term habitats for protozoa

- Low infective dose (10-100 oocysts/cysts)

Molecular diagnostic markers for *Giardia*

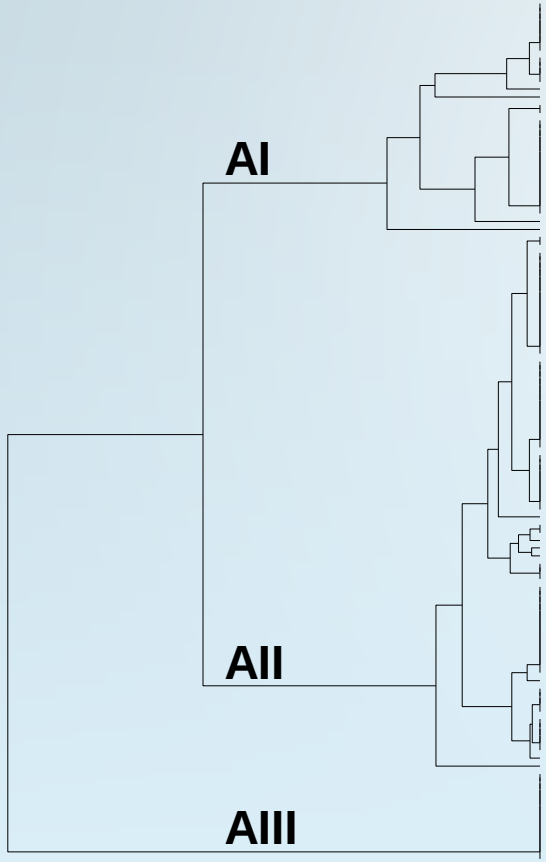
Genotypes and sub-genotypes of Giardia can be identified by analysis of single or multiple loci

For the primary diagnosis of giardiasis
Small subunit ribosomal ribonucleic acid (SSU rRNA)
beta-giardin
triosephosphate isomerase (TPI)
intergenic spacer (IGS) regions

…for further genotyping
SSU rRNA
glutamate dehydrogenase (GDH)
TPI
beta-giardin
IGS region
elongation factor 1-alpha (EF1-alpha)

Large genetic variability within Giardia Assemblages

- molecular analysis limited to single loci suggests potential zoonotic transmission (low resolution level)
- multi-locus sequence typing revealed the existence of sub-groups with defined host associations thus suggesting zoonotic transmission doesn't occur commonly (higher resolution level)



Calf, sheep,cat, deer, human (few cases)

Human, cat (1)

Some species of deer, wild boar, cat

Multilocus data for assemblage A

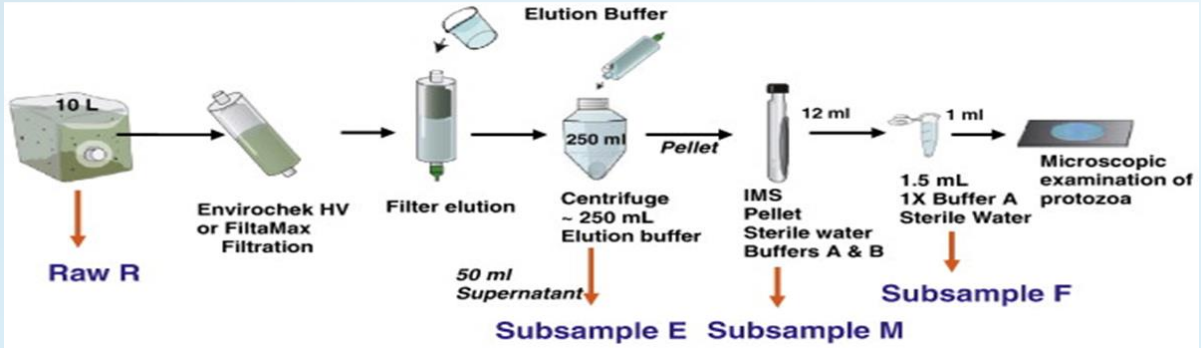
Giardia and water

- Surveys performed in many countries from all over the world.
- Cysts are **ubiquitous** in the aquatic environment
- Detected in wastewater, surface water, ground water, springs and drinking water samples.

Yes, *Giardia* is just everywhere!

Tongariro National park, New Zealand

Filtration **Immunomagnetic separation** **Microscopy**



Health

- Water in streams, rivers and lakes may be unsuitable for drinking due to contamination by Giardia or other bugs. If you doubt the purity of water for drinking you should:
  - boil the water for at least three minutes: or
  - filter the water through an approved filter system: or
  - chemically treat the water.

Standardised methods available
USEPA method 1623
The Water Supply (Water Quality) Regulations 2000 (SI 2000/3184)

SEWAGE RAW WATER SURFACE DRINKING

Country	*Giardia* prevalence in different water samples and cyst numbers detected in water			
	Sewage water	Raw water	Surface and bathing water/swimming pool	Drinking water
Netherlands			58.6%/5.9% Range: 0–167/10 L	
Portugal		15.5%/57.9% Mean: 0.1–108.3/10 L		25.4%
Switzerland			97.5% Range: 0–216/20 L	
Germany		63.8% Range: 0–1314.3/100 L; average 88.2/100 L		14.9% Range: 0–16.8/100 L; average 3.77/100 L
Denmark				
Norway		11.7% Range: 1/10 L		
Finland	Influent 100% Effluent 50%	33.3%	35%	
Sweden				
Greece			29.6% Range: 0–3205/100 L	
Hungary	100% Range: inflow 320–5760/L, outflow 0.6–375/L	48.4%/76.9% Range: 0–1030/100 L	33.3% Range: 0–0.8/L	27.2% Range: 0–63.6/100 L
Czech Republic		Range: 0–485/100L		
France		84.2%/93.8% Range: 0.5–180/10 L	33.3%/67.8%/96.7% Range: 0–511.5/10 L	
Russia		0–357/2 L		
Bulgaria	0–1208/2 L		0–232/2 L	0–255/2 L
Spain	100% Mean influent: 89–8305/L Mean effluent: 79–2469/L	26.9–55.5% Mean: 1–12.8/L	92.3% Mean: 2–400/L; range: 0–722/L	19.2–26.8% Mean: 0.5–4/L
Italy	100% Mean: 60–7000/L	57.1% Range: 0–8/100 L	71% 0.006–80/L	0%
Poland			2–6.9%	0%

Number of positive samples and number of cysts decrease

Wastewater in Europe

Country	Cyst/liter	% positive	method
France	1300–12,000	100	Light microscopy
France	0.23-25,000	100	IF
France	1000–25,000	100	IF
France	330-4300	100	IF+RT-PCR
Italy	2100-42,000	100	IF+PCR
Norway	100–3,600	93	IF
Spain	67	100	IF
Spain	2–14,400	98	IF
Scotland	5-940	96	IF
Scotland	10-3,600	70 – 91	IF

Almost all sample are positive, and very high number of cysts

Surface water in Europe

Country	Sampling	Cyst/l	% positive	method
Finland	39 sites, 139 samples	Not reported	14	IMS+IF+PCR
France	Paris, 5 sites, 57 samples	0-16	66	IMS+IF
Netherlands	Amsterdam, canals Recreational water	0-16 0-1	97 37	IMS+IF
Spain	Rivers, 56 samples Reservoirs, 36 samples	0-250 0-13	92 55	IMS+IF
Switzerland	Recreational water, 2 sites, 40 samples	0-22	97	IMS+IF
Portugal	Surface water, 30 samples Groundwater, 39 samples	Not reported Not reported	57 59	IMS+IF+PCR
Hungary	Rivers, 5 samples Reservoirs, 11 samples	0-3 0-9	100 45	IMS+IF+PCR
Italy	Rivers, 39 samples Watershed, 48 samples	1-465 0-0.1	100 0-50	IMS+IF

Positivity still high, but number of cysts much lower

Genetic identity of *Giardia* cysts in water

Country	Type of water	No of samples	% positive	Range cysts/liter	Assemblage
Hungary	Surface	16	62.5	0.1-17.4	A and B
Portugal	Surface	69	58	-	A
Spain	Surface	116	67.2	2-722	A and E
Norway	Sewage	72	53	100-51333	A and B
France	Wastewater	24	100	530-11000	A and B
France	Wasterwater slaughterhouse	12	58.5 92	1400-4100 -	A E
Italy	Wastewater	16	100	2100-42000	A and B

Worldwide waterborne outbreaks

- **11% of worldwide waterborne outbreaks (1991-2008) due to parasites**

- **>900 water-associated outbreaks** of **parasitic protozoan** disease have been reported until 2016

- **381 waterborne outbreaks between 2011- 2016**: 49% occurred in Australia and New Zealand, 41% in North America, 9% in Europe (Ireland, UK, Norway, Sweden. Belgium, Germany)

- **Cryptosporidium spp. (63%) and Giardia spp. (37%)** were almost exclusively associate with reported outbreaks (Efstratiou et al., 2017).

- **Cryptosporidium outbreaks** are significantly associated with **contamination of recreational water**

- **Giardia outbreaks** are associated with **drinking water** (mainly surface water and community water systems) and **deficiencies in water treatment process** (insufficient barrier and inadequate or poorly operated treatment and disinfection systems)

Accurate estimation of waterborne disease requires the implementation of effective surveillance systems covering all countries!

Epidemiological investigations and successful isolation and identification of the outbreak causative agent from the suspected infection source (water, food) is difficult since:

- sampling occurs days (if not weeks) after the fact!
- causative conditions are rarely continuing and therefore not identifiable

Analysis of patient samples may provide the common infecting agent!

However, clinical manifestations of giardiasis is very variable eventually limiting the identification of infected individuals

Asymptomatic giardiasis: this represents a considerable percentage of infections. The role of carriers in the epidemiology of giardiasis is not well understood

Symptomatic giardiasis: the most common symptom is diarrhea, followed by nausea and abdominal pain. In severe cases, a malabsorption syndrome can occur

Chronic giardiasis: in a small number of cases (15%), the infection tends to be chronic, and resistance to treatment has been observed

Investigation of outbreaks

- Despite the large number of waterborne outbreaks, only **a few have been investigated** to include genetic characterization of cysts from suspected sources and confirmed cases
- What have we learned from those few studies?

Beaver fever and Giardia waterborne outbreaks?

Ann Intern Med. 1980 Feb;92(2 Pt 1):165-70.

Municipal waterborne giardiasis: an epidemilogic investigation.

Beavers implicated as a possible reservoir.

""The **beaver** has gained attention as a potential source of *Giardia* contamination of lakes, reservoirs and streams, but human fecal wastes are probably as important""

"Small aquatic or semi-aquatic wild mammals, such as beavers, muskrats, and small rodents harbor water-born cysts of *Giardia* and **serve as important reservoir hosts**"

Outbreak investigation

- **2000** 12 human samples, 1 sample of filtered water and 1 beaver sample from the outbreak analysed by PCR (SSU-rRNA)
- ✓ Only assemblage A found in 6 human samples, all other samples negative
- ✓ This is the only outbreak caused by assemblage A

No involvement of beavers?
(when positive, beavers harbor assemblage B)

Lesson?

- In the developed world, waterborne transmission is usually the result of contamination with *Giardia* of human origin or a process failure by water utilities, industry or in swimming pools.
- Such contamination may impact negatively on ecosystem health leading to infections in aquatic wildlife which may then establish reservoirs of human infection. The role of the beaver as a '**spill back**' reservoir of *Giardia* in North America is the best known example

And with the advent of Next Generation Sequencing?

Next generation sequencing (NGS) methods or “high through-put sequencing” can generate **millions of sequences per sequencing run** and are increasingly used in the investigation of foodborne outbreaks, particularly for bacterial pathogens (Sekse et al., 2017), but are in their infancy for parasites.

Whole-genome sequencing (WGS) was used to characterise *G. duodenalis* isolates (n = 89) and link Giardia from beavers as the cause of two small community waterborne outbreaks (Tsui et al., 2018)!

Whole genome sequencing (WGS) was applied to characterize *Giardia duodenalis* isolates (N=89) from both outbreak and sporadic infections
29 isolates from raw surface water; 38 from humans; 22 from veterinary sources

Single nucleotide variants and epidemiological data combined!

- Assemblages A (A1 and A2) and B (variables) were identified in surface water, human,and veterinary isolates.
- Assemblage A1 showed little genetic variation among animal and human hosts in isolates collected from across the globe
- Mixes of zoonotic assemblages A and B were seen in all the community waterborne outbreaks in British Columbia (BC), Canada, studied.

Most human infection isolates in waterborne outbreaks clustered with isolates from surface water and beavers implicated to be outbreak sources by public health.

In-depth outbreak analysis demonstrated that **beaver is a possible source of human infections** from contaminated surface water, while acknowledging that theirs is only one role in the complex cycle of zoonotic spread.

AN OUTBREAK OF GIARDIASIS IN ITALY? WHEN? WHERE?

The outbreak occurred in Cambrembo, a very small village located in a mountainous region (1421 meters above sea level) in the North of Italy (Bergamo province).

The village has very few residents, but during winter and summer holidays, up to 100-200 people move there

What happened?

- At the end of 2017, about a hundred people spent Christmas holidays in the village. During early January 2018, the University Hospital of Pavia and the Sacco Hospital of Milan registered an unexpected number of patients with enteric symptoms.
- Overall, 75 individuals reported enteric symptoms, including diarrhoea (97%), fever (41%), nausea (41%), vomiting (37%) and abdominal cramps (24%).
- On January 4, samples from the water supply in Cambrembo were tested and high numbers of thermotolerant coliform bacteria (130 UFC/100 mL), *E. coli (*47 UFC/100 mL) and Enterococci (180 UFC/100 mL) were found.
- Inspection of the water system found that the system pipes were old with signs of leakage.
- The local authority issued a boil water notice. The event attracted the attention of local newspapers


IL GIORNO BERGAMO

CRONACA SPORT COSA FARE EDIZIONI TRUFFA BADANTI FEBBRE DENGUE BIMBA MALTRATTATA SPECIALI

HOME › BERGAMO › CRONACA Pubblicato il 3 maggio 2018

Cambrembo, residenti esasperati: "Da 5 mesi senza acqua potabile"

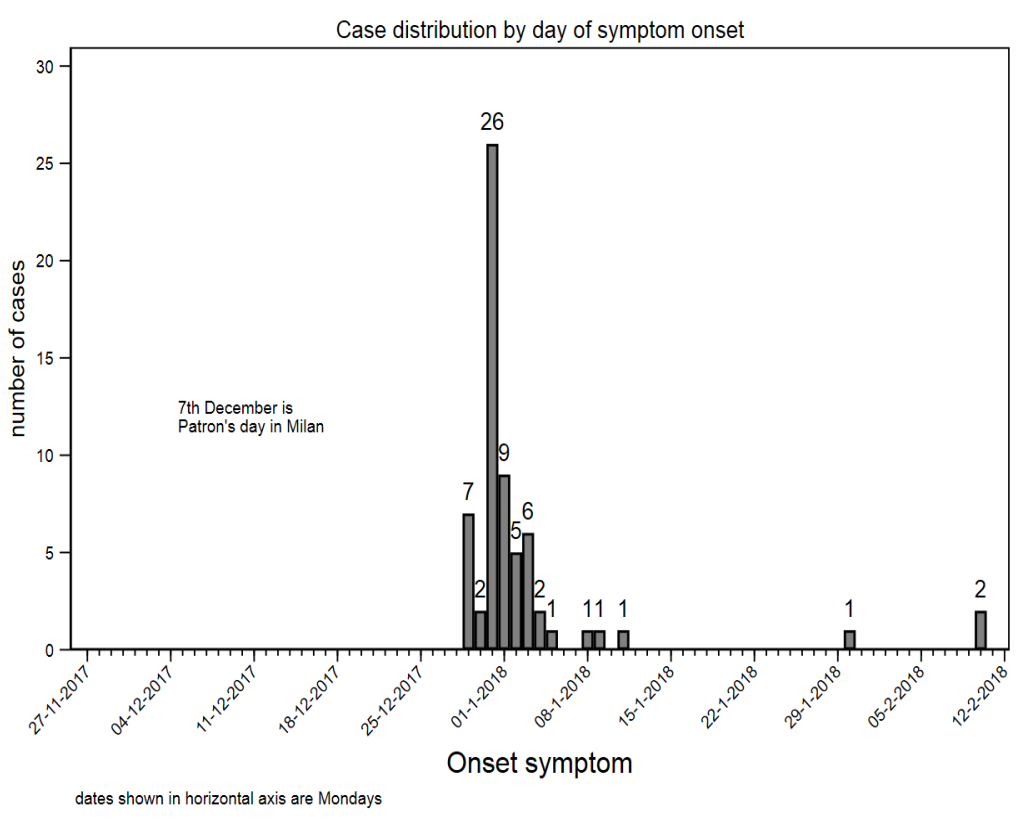
Gli abitanti della frazione di Valleve pronti all'azione legale

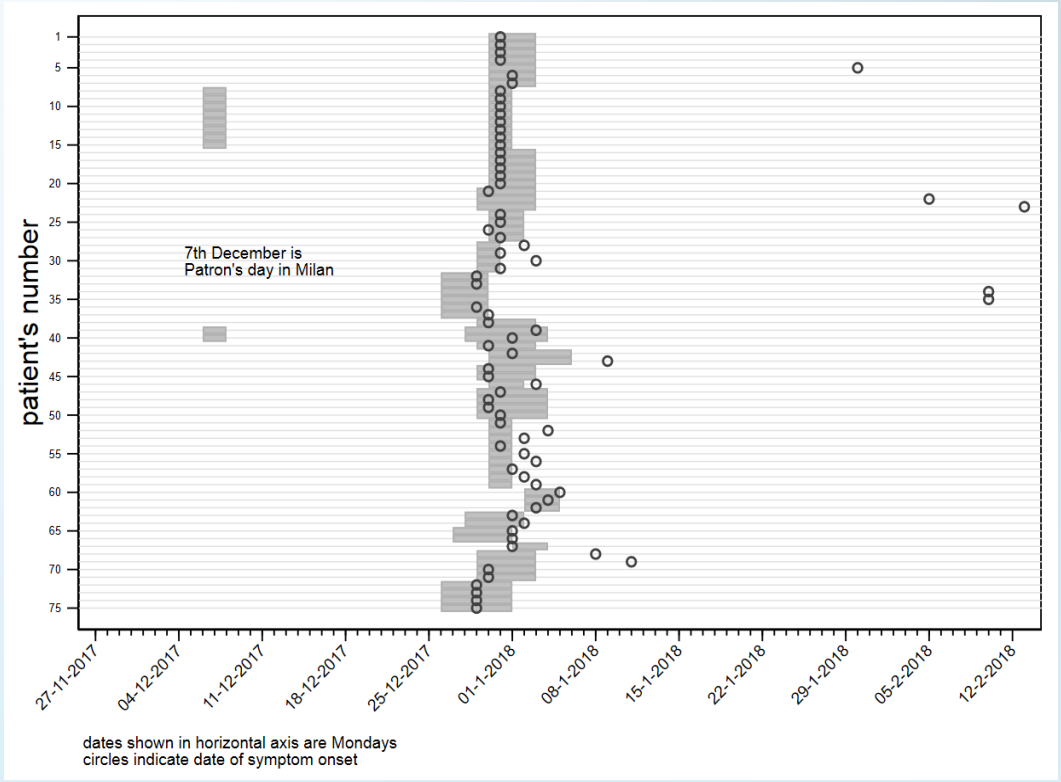

LA VOCE DELLE VALLI

QUOTIDIANO ONLINE DI VALLE BREMBANA E VALLE IMAGNA

Il caso a Cambrembo di Valleve: l'acqua potrebbe non essere la causa dei malori

Epidemiologic data supporting the outbreak





Epidemic curve showing the onset of symptoms.

Grey band showed the time of stay in Cambrembo. Circle indicate date of symptom onset.

Diagnostic investigations

- Diagnostic tests were performed at the two Hospitals on 37 individuals, based on:
  - ✓Microscopic analysis of faecal smears
  - ✓Detection of *Giardia*, *Cryptosporidium* and *Entamoeba* by the Triage Micro Parasite Panel
  - ✓Immunofluorescence for *Giardia* (confirmatory test)
  - ✓Culture methods and isoenzyme typing for *Entamoeba*

- Results showed the presence of *Giardia* and *Entamoeba*

Standard molecular analysis

- In total, **78 stools** from **35 patients**, belonging to **8 families**, were available at ISS for molecular investigations.
- DNA was extracted from the stools and **tested for *Giardia* by nested PCR at the beta-giardin** locus, and for ***Entamoeba* by PCR at the 18S rDNA locus**.
- Samples from 28 patients (**75%**) were **positive for *Giardia*** and sequencing revealed **genotype A2** in all but two patients (one had genotype A3, the other a new variant closely related to A2).
- When multiple samples from the same patient were tested, the same Giardia genotype was detected.
- Samples from 14 patients (38%) were positive for *E. dispar*. No evidence of infection with *E. histolytica* was found.
- **Seven patients were co-infected by *Giardia* and *E. dispar***.

A comparative genomics study

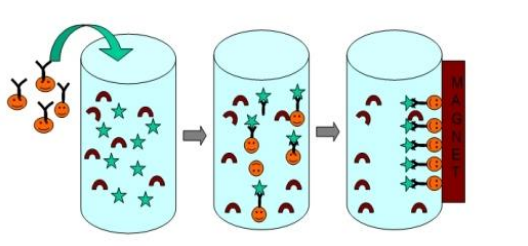
- We were interested in generating genome information from samples collected during the outbreak.

How many (oo)cysts do we need?

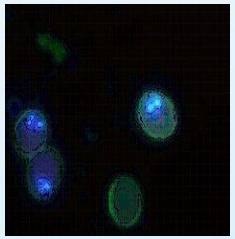
- A single *Giardia* cyst contain about 112 femtograms (10^{-15} g) of genomic DNA
- The minimum amount of DNA for a NGS experiment is 10 nanograms, which in turn corresponds to 25-50,000 cysts.
- Since we cannot grow (easily) these parasite in the lab, what we have is what is in the sample.
- Also consider that any purification step will cause a reduction of the original number of organism. It follows that a sufficient amount of DNA could not be obtained from all samples.

Purification of cysts

This is necessary because cysts are always outnumbered by other organisms present in the sample.



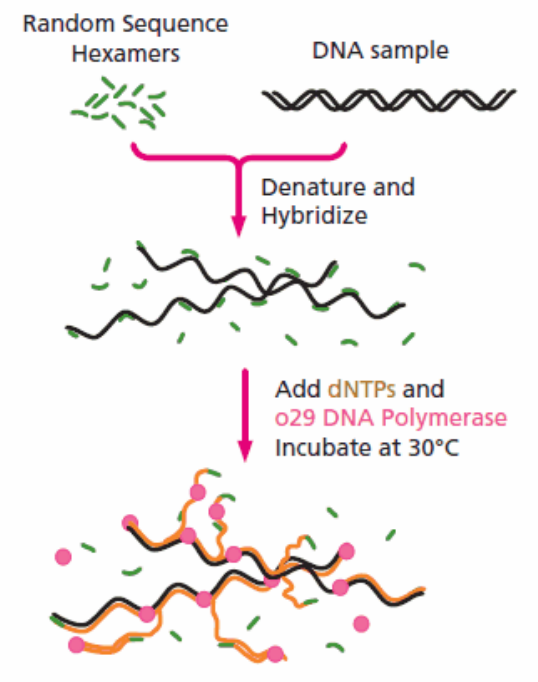
Immunomagnetic separation of cysts from stools

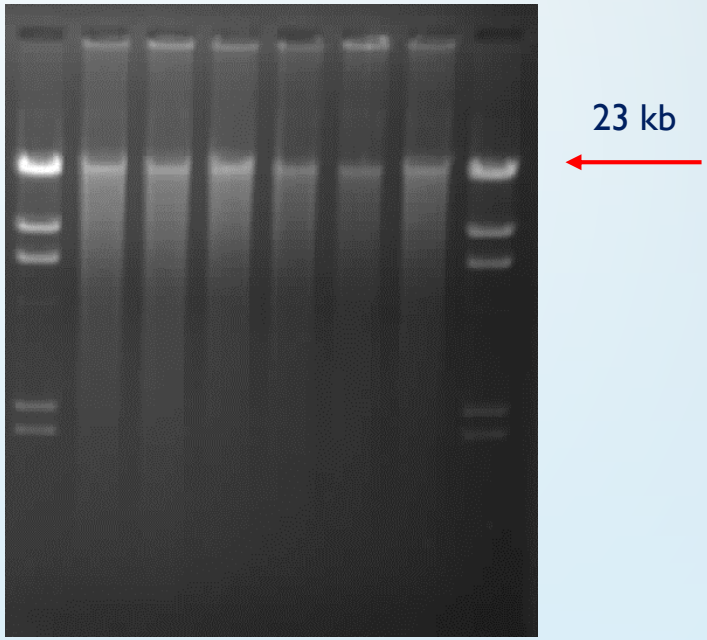


Enumeration of the cysts after double staining with monoclonal antibodies and DAPI

Genomic DNA versus Whole Genome Amplification

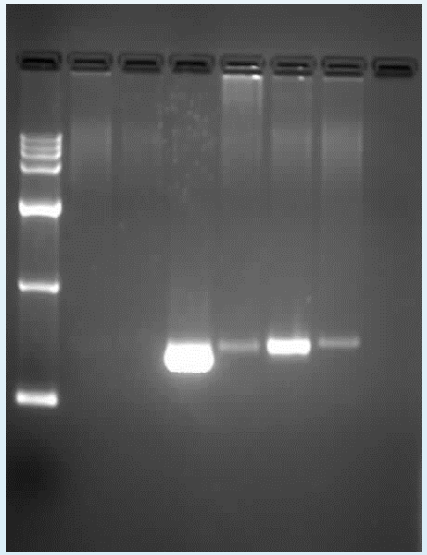
- A way to overcome this limitation is to submit the genomic DNA to Whole Gemome Amplification





Is the DNA «clean» enough?

- It is important to check for enrichment of the target organism, but also to evaluate the presence of possible contaminants.
- We run a 16S-rRNA based PCR to detect the presence of bacterial contamination in the purified (or WGA-amplified) genomic DNA.



Quality checks:
-cloning and random sequencing
-qPCR to detect increase in target DNA

Wet-lab part

- Unfortunately, most stool samples were received in very limited amounts or after long storage in the lack of preservatives.

- Capture of *Giardia* cysts by IMS was attempted on 30 samples from 29 patients. The number of purified cysts varied from zero (7 samples) to 2 millions. In most samples, the number was in the 10,000 range.
-
- After DNA extraction and Whole Genome Amplification, only 12 samples were considered suitable for NGS.

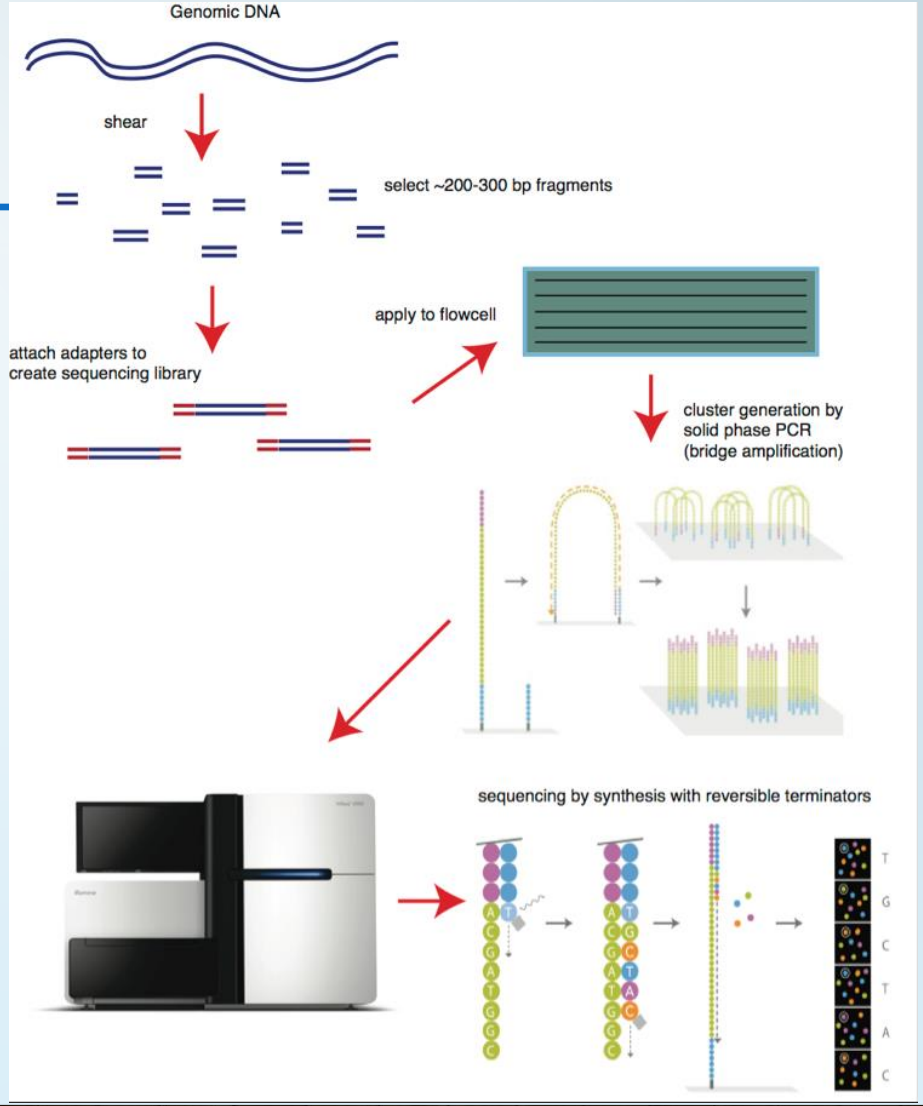
- An illumina Hi-Seq platform was used to sequence libraries (2x150 bp, paired-ends), and to generate 13-19 million reads

Illumina sequencing platform

- In this method, fragmented DNA molecules and primers are first attached on a slide or flow cell and amplified with polymerase so that local clonal DNA colonies, "DNA clusters", are formed.

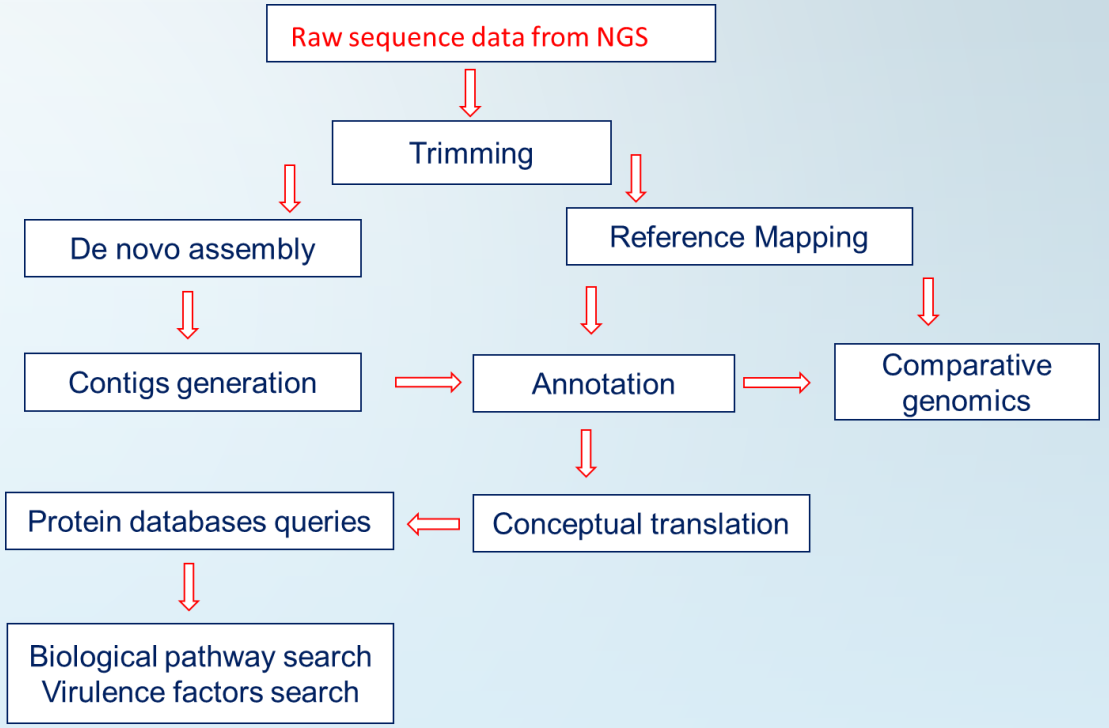
- To determine the sequence, four types of reversible terminator bases (RT-bases) are added and non-incorporated nucleotides are washed away. A camera takes images of the fluorescently labeled nucleotides. Then the dye, along with the terminal 3' blocker, is chemically removed from the DNA, allowing for the next cycle to begin.

- The DNA chains are extended one nucleotide at a time and image acquisition can be performed at a delayed moment, allowing for very large arrays of DNA colonies to be captured by sequential images taken from a single camera.



Genome data analysis

- **After NGS, only 6 isolates generated results** useful for the comparative genome analysis.
- The raw sequence data of these Italian isolates, as well as those from 5 Canadian isolates (Tsui et al., 2018), all belonging to genotype A2, were included in the study.
- Another isolate of genotype 1 (Vanc35, also from Tsui et al., 2018) was included for comparison.
- Raw sequence were mapped against the WB genome (genotype A1).



NGS statistics

SAMPLE ID	ORIGIN AND NAME	N. OF READS	COVERAGE	MAPPED READS (%)
SRR3177753	Canada VANC23	4,367,338	124	93
SRR3177750	Canada VANC3	1,678,926	47	95
SRR3177950	Canada VANC107	2,411,097	71	94
SRR3177990	Canada VANC39	1,786,368	37	82
SRR1957168	Canada VANC42	3,169,345	88	94
GVB37	Italy GVB37	19,850,191	222	95
GVB36	Italy GVB36	19,461,708	40	**17**
GVB11	Italy GVB11	16,718,586	178	91
GVB9	Italy GVB9	19,460,480	221	96
GVB7	Italy GVB7	15,696,755	43	**23**
GVB4	Italy GVB4	13,725,981	109	68
SRR3177768	Canada VANC35 A1	2,074,818	44	98

The number of mapped reads is quite variable, from as low as 17% to as high as 96%, reflecting the presence of contaminants in the IMS-purified samples.

This was partially compensated by the high number of reads generated, which allowed to reach **a minimum coverage of 40x** even for the contaminated samples.

Estimated number of total and private SNPs

Sample ID	N of total SNPs	N of private SNPs
VANC23	164.708	5488
VANC3	155.364	3062
VANC107	164.185	4368
VANC39	155.897	3768
VANC42	164.601	4936
GVB37	163.043	725
GVB36	140.192	202
GVB11	160.319	426
GVB9	162.995	758
GVB7	149.492	379
GVB4	149.098	224
VANC35 (A1)	1154	474

In agreement with previous data (Ankarklev et al., 2015), we found that the **genomes of A2 differ from A1 by about 1-1.5% (140.000-160.000 SNPs)**

The number of SNPs found exclusively in each genome (**private SNPs**) was **3000-5500 in the Canadian** isolates and **200-750 in the Italian** isolates.

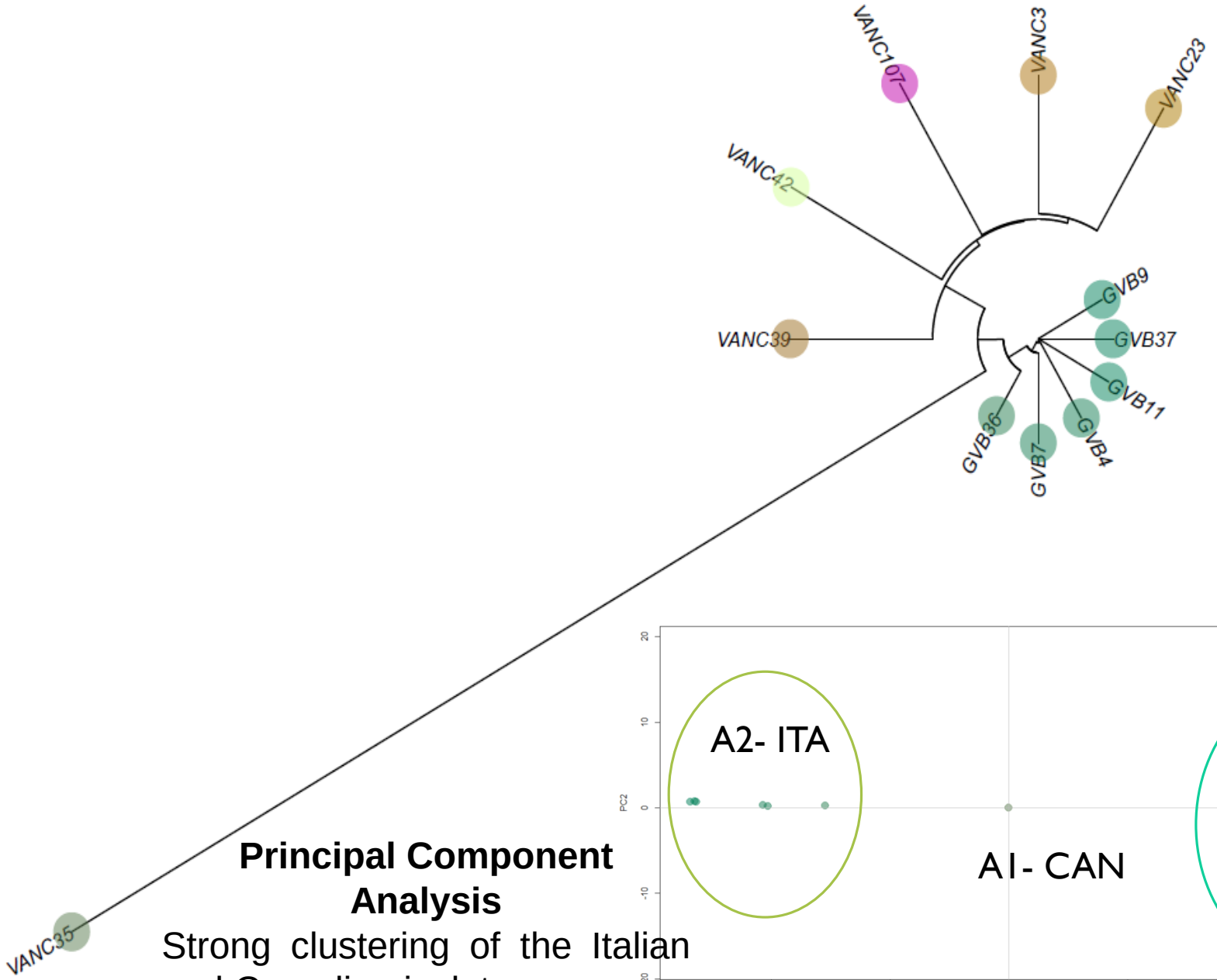
Interestingly, while the total number of SNPs between WB and Vanc 35 (genotype A1) was extremely small, the number of private SNPs (474) was close to that observed in the Italian isolates

Analysis of private SNPs

Sample	% SNPs in non coding	N of VSP with private SNPs	N of NEK Kinase with private SNPs	N of HCM with private SNPs
GVB37	55	30	15	21
GVB36	48	12	4	11
GVB11	46	19	9	15
GVB9	52	29	14	20
GVB7	49	11	11	19
GVB4	53	10	3	13

- About **half** of the private SNPs were found in **non-coding parts** of the genome
- Genes encoding for Variable Surface Proteins, NEK kinases and High Cysteine Membrane proteins contained a large fraction of the private SNPs found in each isolate

Cluster analysis based on all SNPs



Principal Component Analysis

Strong clustering of the Italian and Canadian isolates

Conclusions

- An outbreak of gastroenteritis occurred in a small village in Northern Italy, and involved 75 individuals. Symptoms included diarrhoea (97%), fever (41%), nausea (41%), vomiting (37%) and abdominal cramps (24%).
- The very rapid onset of symptoms suggests the presence of infetious agents other than parasites (e.g., virus)
- **The likely source of infection was drinking water, but efforts to demosntrate the presence of *Giardia* cysts were unsuccesful.**
- **The outbreak was caused by genotype A2, suggesting that water contamination was of human origin**
- This is the **first outbreak of Giardiasis in Italy**, but a second, and larger, outbreak has occurred at the end of 2018 close to Bologna. This was due to Assemblage B and water was again suspected.
- Comparative genomics confirm that the difference between A1 and A2 is 1-1.5%
- **Private SNPs were found in each isolate, so outbreak isolates do not have identical genomes**

ACKNOWLEDGMENT

Istituto Superiore di Sanità, Roma, Italy
Simone M. Cacciò
Anna Rosa Sannella
Paolo Vatta

University of Milan, Italy
V. Zuccaro

San Matteo Hospital, Pavia, Italy
A. Bruno
S Novati
R Maserati

Sacco Hospital, Milan, Italy
R Grande