



EuPathDB.org & ClinEpiDB.org Free Web Resources for Research and Education

Jessica C Kissinger, EuPathDB & ClinEpiDB Consortia







Use Ontologies to manage, integrate and mine rich metadata associated with diverse data types

Many Different categories of molecular data

RNA-SeqFuTranscriptomicsFuArray ProbesPhDNA MicroarrayRTProtein expressionProteinCHIP-SeqImSequence features and motifsKEGenetic variationMaAnnotationSASubcellular localizationMaGenome SequencesTa:Protein targeting andFulocalizationOrChIP-chipStrChEBI CompoundsProtein

Function prediction Function Phenotype (KO, CRISPR) **RT PCR Protein Structure** Immunology **KEGG Metabolic Pathways** Metabolic pathways SAGE **Metabolomics** Taxonomy and Phylogeny **Function** analysis Orthology and synteny Structural variation Protein interactions **RNA Structure**

Compounds and Metabolic Pathway

- 580 unique Data sets (excluding genome sequences)
- Data are provided by the community
- 337 Different genera/species/strains
- 2268 distinct data & metadata terms
 - 152 EuPathDB
 - 125 MicrobiomeDB
 - 2047 ClinEpiDB

Chris Stoeckert – Leads Ontology Team

Provide the power of Boolean operations with the simplicity of Venn diagrams

Home New Search V My Strategies	Add Step 2 : Predicted Signal Peptide Organism 1 selected, out of 30 Filter list below Plasmodium select all clear all expand all collapse all	A EuPathDB Project ene Text Search: synth* Login Register Contact Us mmunity Analyze My Experiment
My Strategies: New Opened	Advanced Parameters	
(Genes) Signal Pep 1079 Genes 5712 Genes Step 1 Step 2	Combine Genes in Step 1 with Genes in Step 2:	Add Step Eact Add Step

Human Computer Interaction studies are very important!

Leverage synteny and orthology to the maximum

- Few genome sequences have annotation and even fewer have curation
- Implement synteny views when possible
- Implement OrthoMCL
- Facilitate inference and data transformations





Users & user communities need to be able to provide input; We need advice about priorities

- Community advisory groups
- Dedicated outreach team (5 people currently)
- User comments
 - Update annotation
 - Link publications
 - Upload images of protein localization, etc.
 - 1043 submitters have provided 12,409 comments relating to 23,281 genes





Different communities need a different UI: ClinEpiDB.org (Same EuPathDB back end)

Main Components

- Tiered data access system as some data are sensitive
- Search Wizard
- Search Strategy System
- Results Tables
- Exploratory Analysis Apps
- Data Download (summary, as approved)

BILL& MELINDA GATES foundation



Good morning from New Dehtil Attending Balastoundation Grand Challenges this meeting to decuse how we can make data more open and

PRISM Cohort Study



Study Details 📀

3 Sites In Uganda from 2011–2017

- The Program for Resistance, Immunology, Surveillance and Modeling of Malaria in Uganda (PRISM) Study is a part of International Centers of Excellence for Malaria Research Program
- Longitudinal cohort and entomological surveillance study designs
- 1,421 Participants from 331 Households with over 48,000 clinical observations

Download Data 📥

SEARCH THE DATA

Ų; 🕋 👬

GEMS1 Case Control



Study Details 🕥

7 S. Asian and African Sites, 2007-2011

- The Global Enteric Multicenter Study (GEMS) investigated the causes, incidence and impact of moderate-to-severe diarrhea in 22,567 children from the Gambia, Mali, Kenya, Mozambique, Pakistan, India and Bangladesh.
- Case-Control study with a 60-day followup visit
- 16S sequence data for ~1000 stool samples available at MicrobiomeDB.org.

Download Data 📩

SEARCH THE DATA



Data Set: PRISM Cohort Study

Summary: Surveillance studies are being conducted in 3 sub-counties in Uganda: Walukuba, Jinja District; Kihihi, Kanungu District, and Nagongera, Tororo District. These areas were purposively chosen to represent varied malaria transmission settings. Walukuba is a relatively low transmission, peri-urban area near Lake Victoria in the south-central part of the country. Kihihi is a rural area with moderate transmission intensity which borders a national park in the south-western part of the country. Nagongera is a rural area with high transmission intensity in the south-eastern part of the country near the border with Kenya. Primary publication: Malaria transmission, infection, and disease at three sites with varied transmission intensity in Uganda: implications for malaria control. Kamya et al. Am. J. Trop. Med. Hyg. 2015;92(5):903-12 Primary contact: Grant Dorsey, UCSF

EuPathDB release # / date: ClinEpiDB 1 / 16-FEB-18

SEARCH THE DATA 👘 🖓 🏦

Expand All Collapse All

1 Annotation, curation and identifiers

Associated Publications

>>

Show how Contents

- Malaria transmission, infection, and disease at three sites with varied transmission intensity in Uganda: implications for malaria control. Kamya et al. Am. J. Trop. Med. Hyg. 2015;92(5):903-12
- Estimating the annual entomological inoculation rate for Plasmodium falciparum transmitted by Anopheles gambiae s.l. using three sampling methods in three sites in Uganda. Kilama et al. Malar. J. 2014 Mar 21;13():111
- Factors associated with malaria parasitemia, anemia and serological responses in a spectrum of epidemiological settings in Uganda. Yeka et al. PLoS ONE 2015;10(3):e0118901
- Mind the gap: house structure and the risk of malaria in Uganda. Wanzirah et al. PLoS ONE 2015;10(1):e0117396
- Measures of Malaria Burden after Long-Lasting Insecticidal Net Distribution and Indoor Residual Spraying at Three Sites in Uganda: A Prospective Observational Study. Katureebe et al. PLoS Med. 2016;13(11):e1002167
- Spatio-temporal analysis of malaria vector density from baseline through intervention in a high transmission setting. Alegana et al. Parasit Vectors 2016 12 12;9(1):637
- Why is malaria associated with poverty? Findings from a cohort study in rural Uganda. Tusting et al. Infect Dis Poverty 2016 Aug 04;5(1):78
- Characterizing microscopic and submicroscopic malaria parasitaemia at three sites with varied transmission intensity in Uganda. Rek et al. Malar. J. 2016 09 15:15():470
- Rapid improvements to rural Ugandan housing and their association with malaria from intense to reduced transmission: a cohort study. Rek et al. Lancet Planet Health 2018;2(2):e83-e94

Contact Grant Dorsey

Contact Institution UCSF

ClinEpiDB 6 7 Mar 2019 ©2019 The EuPathDB Project Team

EuPathDB 🖑 🔘 🧳 🖊 < 🚳 💐 🛞 🛰 👌 🏘







E Assets

III Household wealth index, numerical

Select a Set of Clinical Observations (PRISM) **T**

D Learn about the PRISM Cohort Study



Your Observations filters reduce 18,089 Observations to 3,957

Malaria diagnosis and parasite status 🗙

expand all collapse all
Find a filter Q

Imalaria diagnosis and parasite status

Observation

ITN last night

Visit type

Anthropometry

Diagnosis

III Temperature (C)

III Height (cm)

Weight (kg)

💌 🗮 Malaria diagnosis

Malaria diagnosis and parasite status

Keep checked values at top

Malaria diagnosis and parasite Remaining Observatio... 🕜 🛓 status \$ Observatio... 🕜 Distribution 🕜 % 🕜 18,089 48,721 (100%) (100%) Blood smear indicated but not done 1 (0%) 2 (0%) (50%) Blood smear negative / LAMP negative 5,873 (32%) 17,860 (37%) (33%)Blood smear negative / LAMP not done 3,393 (19%) 11,234 (23%) (30%) Blood smear negative / LAMP positive 1 2,044 (11%) 4,561 (9%) (45%) Blood smear not indicated 1,932 (11%) 5,923 (12%) (33%) 1 Blood smear positive / no malaria 1,913 (11%) 3,126 (61%) (6%) Symptomatic malaria 2,933 6,015 (16%) (12%) (49%)

🔻 🗏 Complicated malaria

Basis of complicated diagnosis

🗏 Severe malaria criteria

E Other diagnosis

1 of 48,722 Observations have no data provided for this filter

Observations shown are for the selected set of participants when they were within the age range specified in the Age filter. Change the default values for the Observations per Year parameter to restrict

observations returned based on your selected observation values.

	Pervations (PRISM) ▼
our <i>Observations</i> filters reduce 18,	,089 Observations to 3,928 Malaria diagnosis and parasite status × Temperature (C) ×
expand all collapse all	Temperature (C)
- Observation	Mean: 36.94 Min: 33.6 Max: 40.2
≣ Malaria diagnosis and parasite status	Select Temperature (C) from 32.2 to 37.99 1 of 48,722 Observations have no data provided for this fil
≣ ITN last night	Q 6000
III Temperature (C)	5000
≣ Visit type	4000
- Anthronomotor	
 Anthropometry 	
Anthropometry Init Height (cm)	3000
	2000 2000
Lill Height (cm)	3000 2000 1000
네 Height (cm) 네 Weight (kg)	2000 1000 0
ાતા Height (cm) ાતા Weight (kg) 👻 Diagnosis	2000 1000
ևև Height (cm) ևև Weight (kg) ✓ Diagnosis ✓ ⅲ Malaria diagnosis	2000 1000 0 33 34 35 36 37 38 39 40 41
Lul Height (cm) Lul Weight (kg) ♥ Diagnosis ♥ III Malaria diagnosis ♥ IIII Complicated malaria	2000 1000 0 33 34 35 36 37 38 39 40 41
Image: Height (cm) Image: Weight (kg) ✓ Diagnosis ✓ Image: Malaria diagnosis ✓ Image: Complicated malaria ✓ Image: Basis of complicated diagnosis	2000 1000 0 33 34 35 36 37 38 39 40 41



©2019 The EuPathDB Project Team





bser	search strategy vations)	atum isin											Stratomu	Clinical Observe	ations(3)*
													Strategy.	ciiriicai Observa	Rename
											Select Columns			×	Duplicate Save As
Asyn	p Parasit	Add Step		r							select all clear al	expand all col	lapse all reset to currer	Q Q	Share
550 (bservations Step 1	Add Step			Age at visit	t (years	s)						×		Delete
	step i				Histogram	Data	Α	symp	tomati	c inc	reases	with a	ge		
_					600			- y p					80		
50	Observations	Revise	Hide search s	trategy panel	1000										
					500	<u> </u>									
bse	rvation Results	Analyze Res	sults							-		-	-		
Firs	t 1 2 3 4 5 Next	Last Adva	nced Paging		400	<u> </u>				°				Add to Basket	Add Columns
					2								-		
	Observation Id	🗢 Visit 🥹 🔒	🗢 Visit 🔕 date	Age at visit (years)	# of Observations								!	Roof material	Wall material
	338619208	Enrollment	2012-08-06 00:00:00.0	0.5	40 Jo 200									Metal	Mud
۵	339419320	Enrollment	2012-11-20 00:00:00.0	0.55			°							Metal	Mud
6	322718875	Enrollment	2011-09-12 00:00:00.0	0.56	100	•							te	Metal	Cement
	339819389	Enrollment	2013-02-05 00:00:00.0	0.57	0	10 2 M	14 2 S	43 43 43		visit (years)	3 48 44 4	2 to 5	24 ng	Thatch	Mud
0	340619540	Enrollment	2013-07-07 00:00:00.0	0.57					Mean 6.885 M		11.01			Metal	Cement
	336418976	Enrollment	2011-12-16 00:00:00.0	0.6	Size of b	ins: 1		0						Metal	Mud
	302718862	Enrollment	2011-08-29 00:00:00.0	0.61	Apply log) ₁₀ scale:	# of Ob	servations						Metal	Mud
۵	344319863	Enrollment	2014-05-19 00:00:00.0	0.66	37.1			LAMP po	sitive	Traditiona	al (all other homes)	Open	Earth or sand	Thatch	Mud
		Enrollment	2013-06-27	0.74	36.8			Blood sm LAMP po	ear negative /	Traditiona	al (all other homes)	Open	Earth and dung	Thatch	Mud
	340519540	LINOIMEN	00:00:00.0					LAMP po	31140						

Hide search strategy panel	
(Participants)	Make it easy:
Asymp Parasit 3550 Observations Step 1 Step 2 Add Step	View Longitudinal Data for Each Participant

rticipar	t Results Analyz	e Results								
irst 1	2 3 4 5 Next Last	Advanced Paging	1						Download Add to Bask	Add Columns
ipant	≑ Age at enrollment 🎱 📠	Household ID	Dwelling 3 type	+ Household wealth index, numerical	≑ Malari	ia status timeline 🎱				
	0.5	HH119008702	Traditional (all other homes)	-0.93	121	2012	0 00 00 0 00 0 2013	2014	2015	2016
	0.5	HH116000604	Traditional (all other homes)	-2.76	191	2012	D (1) (9) (9) (9) (9) (9) (9) (9) (9) (9) (9	2014	0 0 0 0 0 00 000000 2015	2016
	0.5	HH118002402	Traditional (all other homes)	-1.38	1991 000	2012	200 C CONCERCIÓN CON 2013		2015	000 000000 2016
	0.5	HH132012601	Traditional (all other homes)	-2.42	1991 899	2012	2013	2014	2015	0 0 0 C
	0.51	HH101009801	Traditional (all other homes)	0.76	1991 899	2012	200 00 00 00 00 00 00 00 00 00 00 00 00	2014	0000 • 000 •	2016
	0.51	HH120005401	Traditional (all other homes)	1.05	200 H	2012	2013	2014	2015	2016
	0.51	HH124003601	Modern (wood, cement or brick walls; metal or tiled roof and closed eaves)	0.0086845253	19 88	2012	2013	2014	2015	2016
	0.52	HH107020201	Traditional (all other homes)	-0.93	131	2012	€C (C) (C) (C) (C) (C) (C) (C) (C) (C) (C	2014	2015	0 C
	0.52	HH118007201	Traditional (all other homes)	-0.81	181	2012	2013	2014	2015	2016





Bioinformatics Resource Center Community Evolution



Our User Base is Primarily Located Tropical Disease Endemic Countries



A highly-utilized resource that is not well known outside of the pathogen communities







Project Pls: David Roos – UPENN Jessie Kissinger – UGA Christiane Hertz-Fowler - Liverpool



Our goal: enabling end users in the lab, field & clinic to make effective and appropriate use of large-scale datasets, expediting discovery research and translational application by making data FAIR