# Oligotyping EF-1a amplicons:

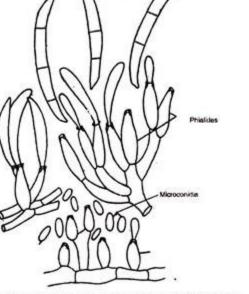
A method for subspecies investigations of Fusarium oxysporum

9-8-17

Peter Henry

PhD Student

Pls: Tom Gordon and Johan Leveau Dept. Plant Pathology at UC Davis



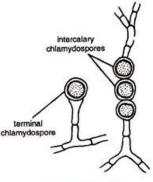


Fig. 4. Fusarium : Chlamydospores

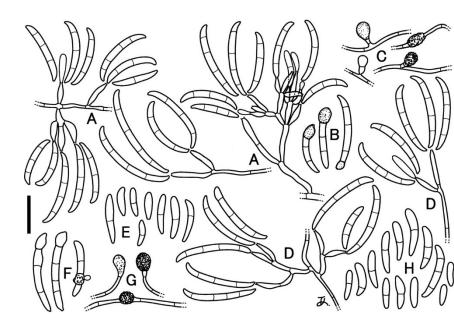
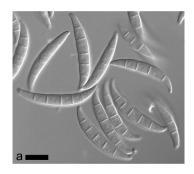
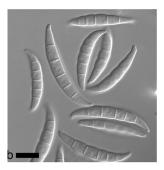


Fig. 3. Fusarium : Phialide, microconidia and macroconidia.



### Overview



#### 1. Introduction to my study system

- 1. Fusarium
- 2. Background on the translation elongation factor 1-alpha locus
- 2. Explanation of oligotyping: getting more from your amplicons
  - 1. Shannon's Entropy
  - 2. Oligotyping
- 3. Pipeline & performance on mock communities
  - 1. vsearch **inter**specific relative abundance
  - 2. Oligotyping **intra**specific relative abundance
- 4. Where you can learn more

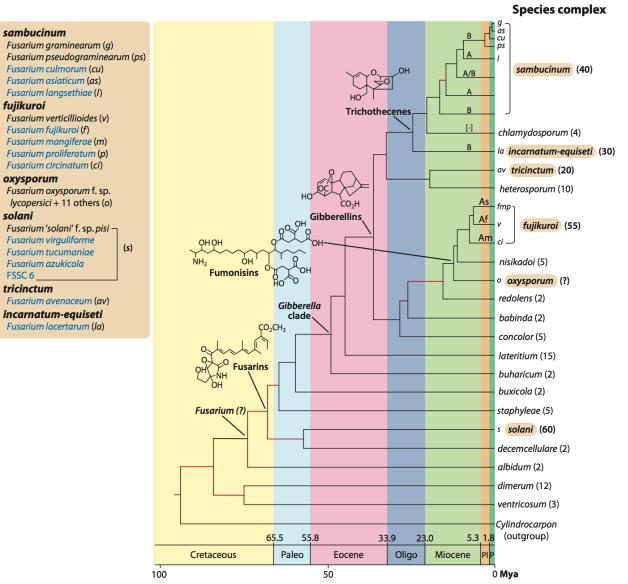
#### Fusarium

fujikuroi

solani

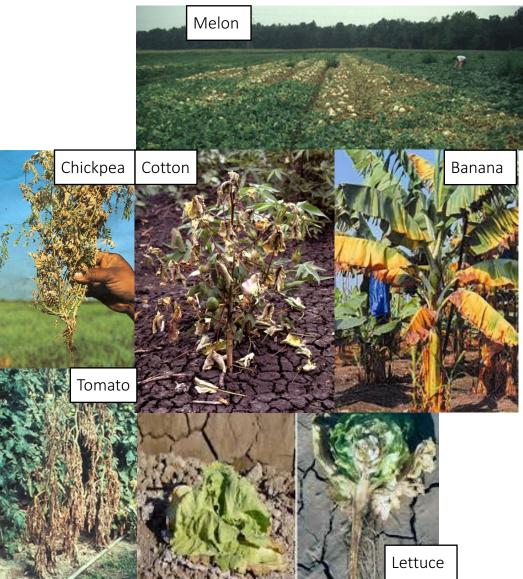
FSSC 6

- A genus of Fungi containing:
  - Many economically important plant pathogens
  - Some human pathogens
  - Soil/plant-associated species



#### Fusarium oxysporum

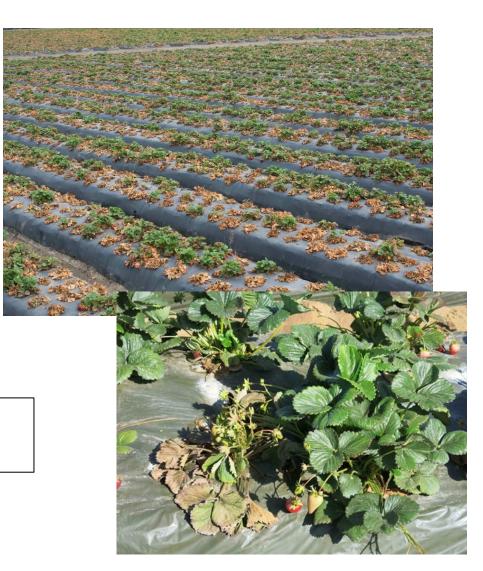
- Entire species = broad host range Over 70 reported hosts, primarily vegetable crops
- Individual strains = narrow host range Phenotypic characterization based on preferred host: "formae speciales"
- Characteristics:
  - Ubiquitous in soils across the globe
  - Highly diverse:
    - Most strains are non-pathogenic
    - Non-pathogenic strains can have disease suppressive effects
  - Main niche: root endophyte, primary consumer of dead plant tissue

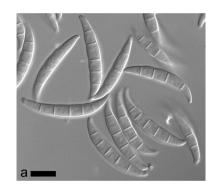


#### Fusarium oxysporum

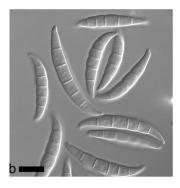
- What is the extent of diversity in soils?
- What environments are more diverse?
- What influences persistence in soil?
- How can we manage agricultural environments to decrease persistence of pathogenic strains?

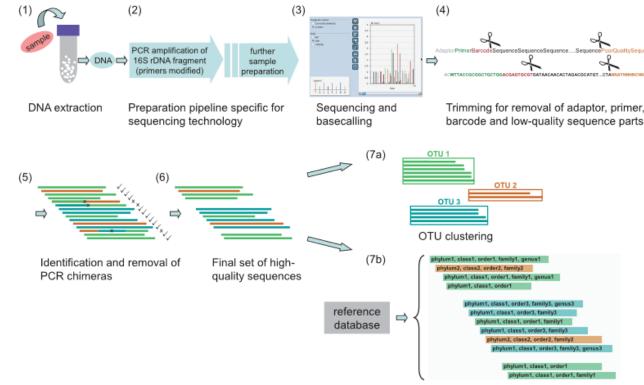
We need a high-throughput tool to assess changes in relative abundance of *Fusarium oxysporum* strains...





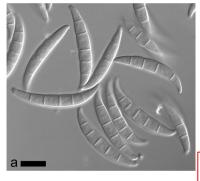
#### Introduction



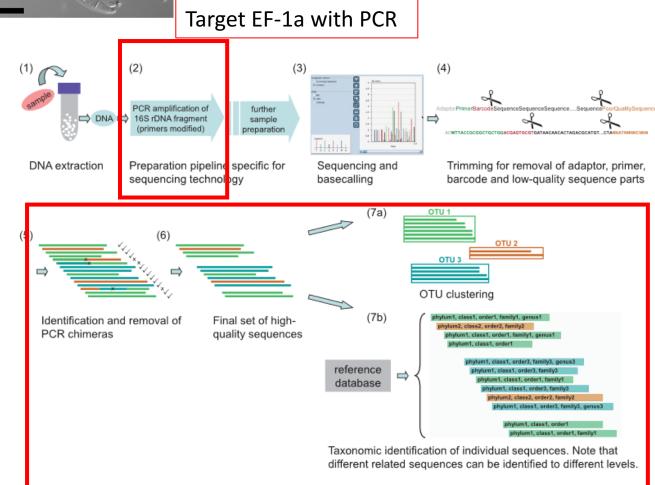


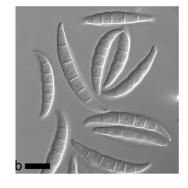
Taxonomic identification of individual sequences. Note that different related sequences can be identified to different levels.

- I want to monitor changes in relative abundance of *Fusarium oxysporum* sequence types
- Typical microbiome analyses of fungi cannot evaluate sub-species diversity



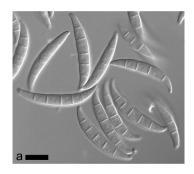
#### Introduction



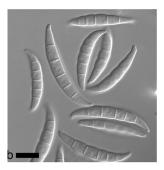


- I want to monitor changes in relative abundance of *Fusarium oxysporum* sequence types
- Typical microbiome analyses of fungi cannot evaluate sub-species diversity

Customize data analysis for this locus



### Overview

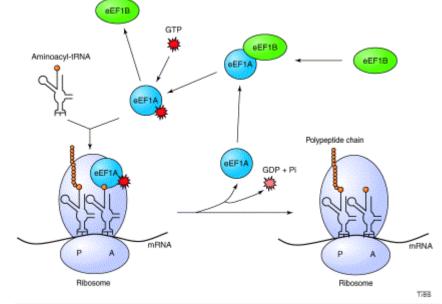


- 1. Introduction to my study system
  - 1. Fusarium
  - 2. Background on the translation elongation factor 1-alpha locus
- 2. Explanation of oligotyping: getting more from your amplicons
  - 1. Shannon's Entropy
  - 2. Oligotyping
- 3. Pipeline & performance on mock communities
  - 1. vsearch **inter**specific relative abundance
  - 2. Oligotyping **intra**specific relative abundance
- 4. Where you can learn more

## Translation Elongation factor 1-alpha



- Conserved, single copy gene in eukaryotes
- Eukaryotes have two elongation factor genes.
- Elongation factor 1 has two subunits: alpha and beta-gamma
  - Elongation factor 1 alpha regulates entry of tRNAs into an available site on a ribosome
  - Elongation factor beta-gamma facilitates release of GDP from the alpha subunit



DOI: <u>http://dx.doi.org/10.1016/j.tibs.2003.11.006</u>

#### History of EF-1a use in phylogenetics

Proc. Natl. Acad. Sci. USA Vol. 93, pp. 7749–7754, July 1996 Evolution

#### The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny

SANDRA L. BALDAUF\*<sup>†</sup>, JEFFREY D. PALMER<sup>‡</sup>, AND W. FORD DOOLITTLE\*

\*Canadian Institute for Advanced Research and Department of Biochemistry, Dalhousie University, Halifax, NS B3H 4H7, Canada; and <sup>‡</sup>Department of Biology, Indiana University, Bloomington, IN 47405

Proc. Natl. Acad. Sci. USA Vol. 90, pp. 11558–11562, December 1993 Evolution

#### Animals and fungi are each other's closest relatives: Congruent evidence from multiple proteins

SANDRA L. BALDAUF\* AND JEFFREY D. PALMER<sup>†</sup>

\*Institute for Marine Biosciences, National Research Council of Canada, Halifax, NS, Canada, B3H 3Z1; and <sup>†</sup>Department of Biology, Indiana University, Bloomington, IN 47405

#### Α

An

Fu

PI

Pr

Ar

		1	z am			111261	CIUI				
_	202			*	* *	0				246	
[X1	GDNMLE	PSPNMI	WFKG	WKIT	RKEGS	GSGTI	LLE	ALDCI	LPP	SRP	
Dm	GDNMLE										
Ov	GDNMLE										
Sc	GDNMIE										
Mr	GDNMLD										
At											
Те	GDNMIE	RSTNLI	DWYKG			PI	LLE	ALDQI	NEP	KRP	
Eg	GDNMIE										
Dd	GDNMLE										
Eh	GDNMIE										
G1	GDNIME										
Ss	GDNVTH										
Hh	GDNIAE	ESEHT	GWYDG			E1	LLE	ALNEL	PAP	EPP	
-	***	0	*o *				0	00 0	*	*	

12 amino acid insertion

#### History of EF-1a use in phylogenetics

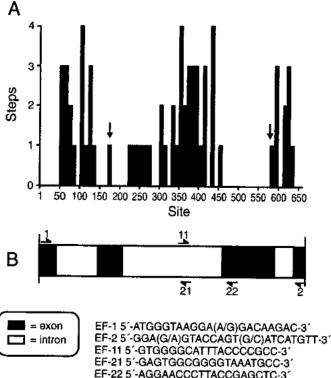
Proc. Natl. Acad. Sci. USA Vol. 95, pp. 2044–2049, March 1998 Applied Biological Sciences

SW 280th Street, Homestead, FL 33031-3314.

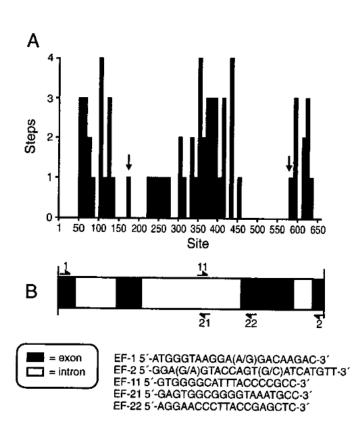
#### disease of banana: Concordant evidence from nuclear and mitochondrial gene genealogies KERRY O'DONNELL\*<sup>†‡</sup>, H. CORBY KISTLER<sup>†§</sup>, ELIZABETH CIGELNIK\*, AND RANDY C. PLOETZ<sup>¶</sup> \*National Center for Agricultural Utilization Research, U.S. Department of Agriculture–Agricultural Research Service, 1815 North University Street, Peoria, IL 61604; <sup>§</sup>Department of Plant Pathology, University of Florida, Gainesville, FL 32611; and <sup>®</sup>University of Florida, Tropical Research and Education Center, 18905

Multiple evolutionary origins of the fungus causing Panama

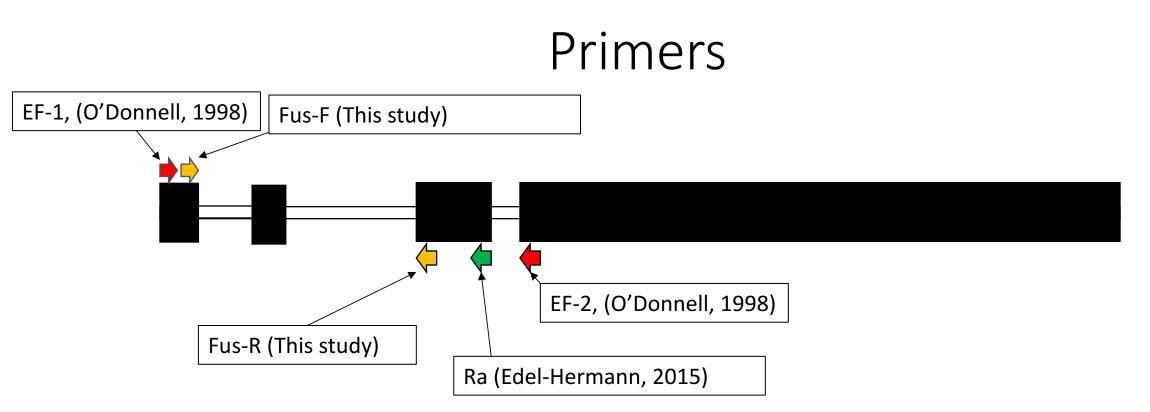
31). Results from the present study demonstrate that the EF-1 $\alpha$  gene, with 95% of the signal derived from intron sequences, possesses 50% more phylogenetic information than the mtSSU rDNA. In sharp contrast to the exons, which lack indels and possess only two phylogenetically informative sites within the 36 taxon matrix, EF-1 $\alpha$  introns appear to be under relaxed evolutionary constraints as inferred from the substitutional pattern (Fig. 1A) that includes indels. However, the low level of homoplasy observed, coupled with the concordance of EF-1 $\alpha$  and mtSSU rDNA gene trees, indicate that informative sites within the EF-1 $\alpha$  gene are not saturated.



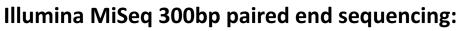
### History of EF-1a use in phylogenetics

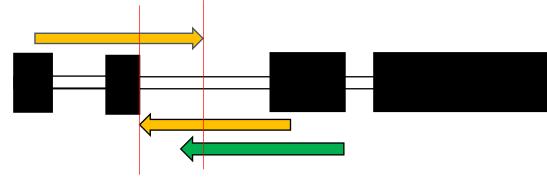


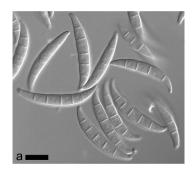
- Subsequent to O'Donnell et al. (1998), EF-1a became widely adopted for phylogenetic comparisons in *Fusarium*
- 2,725 sequences in my custom database:
  - 1,651 from Fusarium-ID database
  - 449 from NCBI Genbank
  - 625 from 14 recent publications
- Final database = 718 unique EF-1a sequences



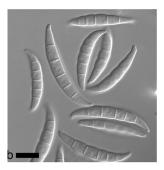
- Fa/Ra amplicon sizes (including primers):
  - *Fusarium oxysporum*: ~572bp
  - Fusarium solani: ~612 bp
- Fus-F/Fus-R amplicon sizes (including primers):
  - Fusarium oxysporum: ~450bp
  - Fusarium solani: ~495bp







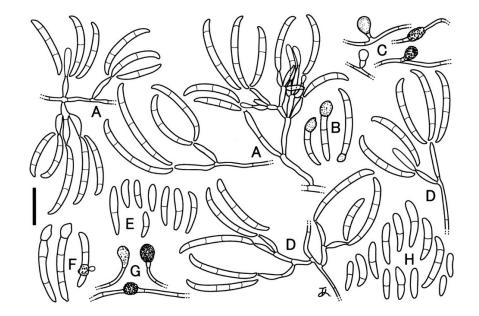
### Overview



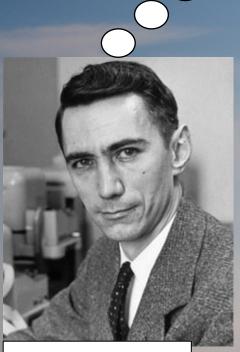
- 1. Introduction to my study system
  - 1. Fusarium
  - 2. Background on the translation elongation factor 1-alpha locus
- 2. Explanation of oligotyping: getting more from your amplicons
  - 1. Shannon's Entropy
  - 2. Oligotyping
- 3. Pipeline & performance on mock communities
  - 1. vsearch interspecific relative abundance
  - 2. Oligotyping **intra**specific relative abundance
- 4. Where you can learn more

#### Takeaway ideas

- Oligotyping can extract additional information from your amplicons
  - Analyzes positional entropy in aligned reads
- It is first necessary to classify reads
  - dbcAmplicons
  - Qiime
  - Mothur
- Reads corresponding to a single species or genus can be aligned and oligotyped



#### what is the relationship between predictability and information:

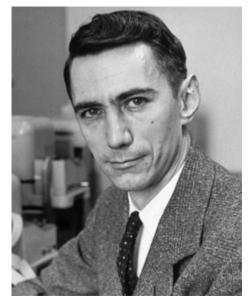


Claude Shannon

## Shannon's entropy: Uncertainty has greater information value

$$H(x) = \sum_{i=1}^{n} -P(x_i) * \log_2(P(x_i))$$

- H(x) = total entropy
- P(x) probability of "x"
- Example: A fair coin (50% heads, 50% tails)
  - f(x) for heads (or tails), = 0.5
    - Heads:  $-0.5*\log_2(0.5) = -0.5*-1 = 0.5$
    - Tails: -0.5\*log<sub>2</sub>(0.5) = -0.5\*-1 = 0.5
  - H(x) = 0.5 + 0.5 = 1



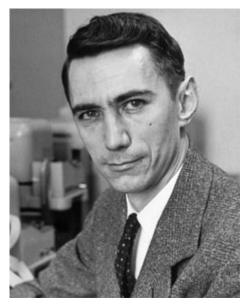
Claude Shannon

• Equal probabilities for all possibilities have the highest Shannon entropy

## Shannon's entropy: Uncertainty has greater information value

$$H(x) = \sum_{i=1}^{n} -P(x_i) * \log_2(P(x_i))$$

- H(x) = total entropy
- P(x) probability of "x"
- Example: An unfair coin (20% heads, 80% tails)
  - Heads: -0.2\*log<sub>2</sub>(0.2) = 0.46
  - Tails: -0.8\*log<sub>2</sub>(0.8) = 0.25
  - H(x) = 0.46 + 0.25 = 0.71
  - Unequal probabilities have lower Shannon entropy



Claude Shannon

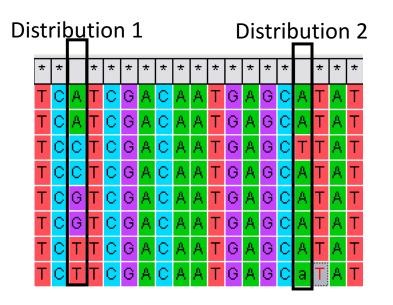
## Shannon's entropy:

## Uncertainty has greater information value

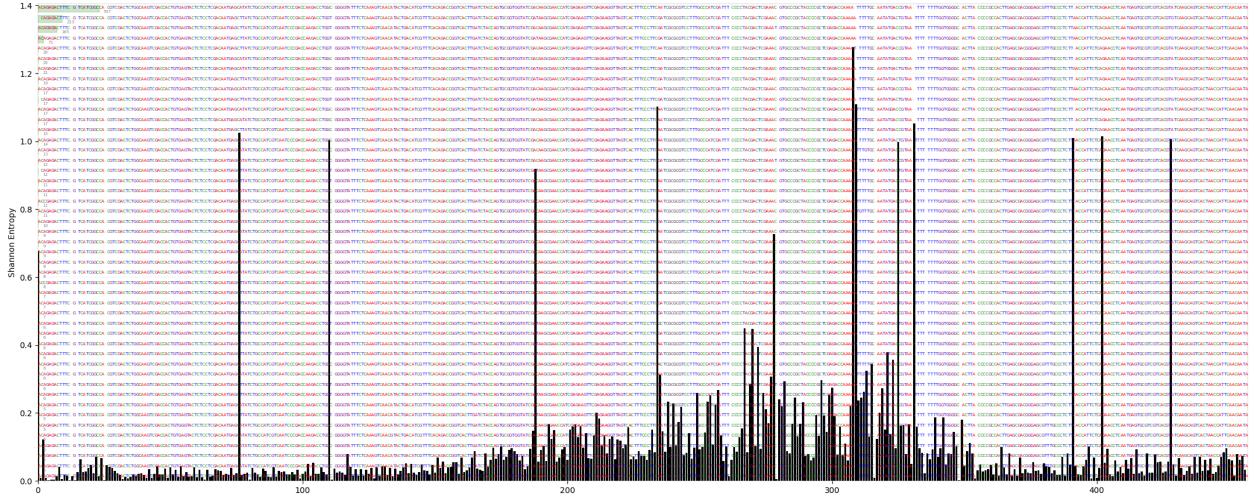
$$H(x) = \sum_{i=1}^{n} -P(x_i) * \log_2(P(x_i))$$

- What about DNA?
- Distribution 1:
  - $-0.25*\log_2(0.25) = 0.5$
  - $-0.25*\log_2(0.25) = 0.5$
  - $-0.25*\log_2(0.25) = 0.5$
  - $-0.25*\log_2(0.25) = 0.5$
  - H(x) = 2

- Distribution 2:
  - $-0.875*\log_2(0.875) = 0.17$
  - $-0.125*\log_2(0.125) = 0.38$
  - H(x) = 0.55
- Distribution 3 (not shown):
  - $-0.998*\log_2(0.998) = 0.003$
  - $-0.002*\log_2(0.002) = 0.017$
  - H(x) = 0.02



#### Shannon's entropy for every position in an alignment



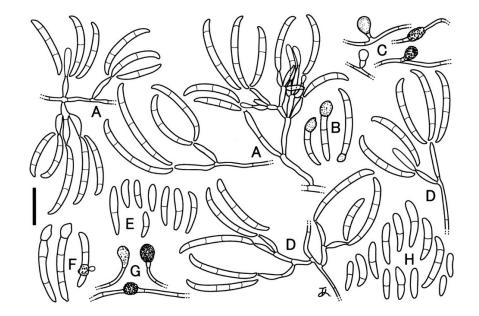
## The Entropy to Oligotype continuum

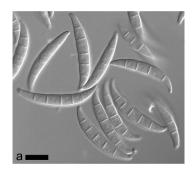
- 1. Determine entropy for every position
- 2. Compare nucleotides at entropic positions for every read
  - 1. CTA = 66%
  - 2. CGG = 20%
  - 3. AGA = 13%
- Min. abundance threshold
- AGG = 0.5%
  CGA = 0.5%
- 3. Set minimum abundance threshold
- 4. Estimate relative abundances of oligotypes
- 5. Reconstruct sequences for each oligotype

	C or		or G	 G or				
0.0	0 0	100 IOO	, no second in 2000 a second i Second i Second i second	••••••••••••••••••••••••••••••••••••••				
	CAGAGACTITE: O TCATCORECA: COTCORCTETOREAAGTESACACTOTOGOSACTETCETCEGACATSAO SCAGAGACTITE: O TCATCORECA: COTCORCTETOREAAGTCSACCACTOTOGOSTACTETCETCORCAATSAO NAGAGACTITE: O TCATCORECA: COTCORCTETOREAAGTCSACCACTGTGAGTACTETCETCGACAATSAO	NTATCTGCCATCGTCAATCCCGACCAAGACCTG NTATCTGCCATCGTCAATCCCGACCAAGACCTG NTATCTGCCATCGTCAATCCCGACCAAGACCTG		OGTCACTTGATCTACCAG TOCOGTGGTATCO CAAG ISGTCACTTGATCTACCAG TOCOGTGGTATCH CAAG ISGTCACTTGATCTACCAG TOCOGTGGTATCH CAAG				
0.2	ACRARACTITE: & TEATORIEGA: OBTEGRETETOREANDTEGREACTETOREANDTEGREATING 5 ACRARACTITE: & TEATORIEGA: OBTEGRETETOREANDTEGREACTETOREANDTEGREATING	ATATCTGCCATCGTCAATCCCGACCAAGACCTG ITATCTGCCATCGTCAATCCCGACCAAGACCTG	C GGGGTATTTCTCAAAGTCAACATACTGACATCGTTTCACAGACC F. GGGTATTTCTCAAAGTCAACATACTGACATCGTTTCACAGACC	OGTCACTTGATCTAIC AG TGCGGTGGTATCG DAG IGGTCACTTGATCTAIC AG TGCGGTGGTATCG TAAG				
	6 ACREMENTITE: 6 TEATESIECA: COTEGACTETISCAAGTESACEACTOTISAGTACTETECTESACAATSAG ACREMENTITE: 6 TEATESIECA: COTEGACTETISCAAGTESACEACTOTISAGTACTETECTECGACAATSAG 6	ITATCTSCCATCGTCAATCCCGACCAAGACCTG ITATCTGCCATCGTCAATCCCGACCAAGACCTG	F GEOGRATTITCTCAAAGTCAACA TACTGACATCSTTTCACAGACC F GEOGRATTTCTCAAAGTCAACA TACTGACATCSTTTCACAGACC	ISGTCACTEGATCTACCAG TGCGGTGGTATCG CAAG ISGTCACTEGATCTACCAG TGCGGTGGTATCG TAAG				
	0 CKGRAGACTITC: G TCA TOSECA: COTCORCTCTOCEAAGTCGACCACTGTGAGTACTCTCCCCGACAATGAG CGGRAGACTITC: G TCA TOSECA: COTCORCTCTOCEAAGTCGACCACTGTGAGTACTCTCCCCGACAATGAG	ATATCTGCCATCGTCAATCCCGACCAAGACCTG ATATCTGCCATCGTCAATCCCGACCAAGACCTG		DIGTCACTIGATCTACEAG TOCOGTOGTATCO CAAG				
0.4	0 ACAGAGACTITE: 5 TEATESEECA: COTCOACTEGECAAGTCGACEACTGTGAGTACTETECTCGACAATAGO COGGAGACTITE: 5 TEATESEECA: COTCGACTCTGCCAAGTCGACEACTGTGAGTACTETECTCGACAATAGO	ITATCTGCCATCGTCAATCCCGACCAAGACCTG ITATCTGCCATCGTCAATCCCGACCAAGACCTG	F. GEGGTA TT TCTCAAAGTCAACA TACTGACATCG TT TCACAGAC F. GEGGTA TT TCTCAAAGTCAACA TACTGACATCG TT TCACAGAC(	INGTCACTIGATCTALEAG IGCOGTGGTATCO TAAG				
				DISTCACTIGATCTACEAS TOCOSTOGTATCO FAAG				
		ITATCTGCCATCGTCAATCCCGACCAAGACCTG ITATCTGCCATCGTCAATCCCGACCAAGACCTG	F. GEGGTA TITICTICA AAGTICA ACA TACTIGA CATOG TITICA CAGACI F. GEGGTA TITICTICA AAGTICA ACA TACTIGA CATOG TITICA CAGACI	OGTCACTTGATCTACCAG TSC GSTGGTATC OF TAAG				
2.5	ACCIGNACTITE: G TEATCORECA: COTEGACTETOREARDICGACEAETOTGAGTACTETEETECAEAATGAG 8 ACCARAGACTITE: G TEATCORECA: COTEGACTAETOGEAAGTEGACEAETOTGAGTAETEETECETEGACAATGAG 7	ITATCTGCCATCGTCAATCCCGACCAAGACCTG ITATCTGCCATCGTCAATCCCGACCAAGACCTG	F. GEGETA TT TCTCAAAGTCAACA TACTGACATOS TT TCACAGAC F. GEGETA TT TCTCAAAGTCAACA TACTGACATOS TT TCACAGAC	OGTORCTTOATCTACCAG TGCOGTOGTATCO TAAG OGTORCTTGATCTACCAG TGCOGTOGTATCO CAAG				
0.6	B	TATCTOCCATCGTCAATCCCGACCAAGACCTG	C GEOGRA TT TCTCAAAGTCAACA TACTGACATCG TT TCACAGACC	OGTCACTTGATCTACCAG TOCOGTOGTATCG				
	CARAGACTITE: G TEATORCEA: OSTEGACTETOSCAASTEGACEACTOTGAGTACTETECTCACAATGAG			IGGTCACTTGATCTACCAGTGCGTGGTATCGCAAG				
0.8	ACAGAGACTITE: G TEATCORECA: COTEGACTACTOCEAAGTCGACEACTGTGAGTACTETECTCCACAATGAG 9 ACAGAGACTITE: G TEATCORECA: COTEGACTETGEAAGTCGACEACTGTGAGTACTETECTCCACAATGAG	ATATCTGCCATCGTCAATCCCGACCAAGACCTG ATATCTGCCATCGTCAATCCCGACCAAGACCTG	E OSOGTA TETETCAAAGTEAACA TACTOACATOS TETEACAOACO E OSOGTA TETETCAAAGTEAACA TACTOACATOS TETEACAOACO	OGTOACTTGATCTACCAG TOCODTGGTATCOCCAGO				
0.0	21 ACMAGAGACTITE & TEATEOREEA ESTEGACTETGERAAGTEGACEACTGTGAGTACTETECTEGACAATGAG	ATATCTGCCATCGTCAATCCCGAGCAAGACCTG	C GOOGTA TT TCTCAAAGTCAACA TACTGACATCG TT TCACAGACC	OGTCACTTGATCTACCAG TOCOGTOGTATCO CAAG				
0.8	ACEGAGACTITE & TEATEGREEA ESTEGACTETGREAAGTEGACEACTGTGAGTACTETECTEGACAATGAG	ATATCTGCCATCGTCAATCCCGACCAAGACCTG	C GEGETA TETECTOAAAGTOAACA TACTEACATOS TETECACAGACO	IGGTCACTTGATCTACCAG TGCGGTGGTATCGCAAG				
	ADRAMAGETTE: G TEATGOREA COTEGAETETOREARGTEGACEAETOTGAGRACTETECETCOREAATTAD 11 ACRAMAGETTE: G TEATGOREA COTEGAETETOREARGTEGACEAETOTGAGRACTETEETECACAATTAD 11	TATCTGCGATCGTCGATCGCGACCAAGACCTG	C GEGSTA TTTCTCAAAGTCAACA TACTGACATCETTTCACAGAC	OGTCACTTGATCTACCAG ISCOSTGGTATCG				
	CAGAGACTITE & TEATOGOECA OFTEGACITETGOEAAGTEGACCACTGTGAGTACTETEETCGACAATGAG			OGTCACTTGATCTACCAG IGCOGTGGTATCG TAAG				
		TATCTGCCATCGTCAATCCCGACCAAGACCTG	C GGGTA TT TCTCAAAGTCAACA TACTGACATOG TT TCACAGAC	IGTCACTIGATCTACCAG IGCOSTIGGTATCGACAAG				
1.0	ACKAGACTITE & TEATEORECA. COTCOACTEOREAGTCOACTACTOTGAGTACTETECTCOACAATGAG 34 ACKAGACTITE & TEATEORECA. COTCOACTEOREAGTCOACTACTOTGAGTACTETECTCOACAATGAG			OGTCACTTGATCTACCAG TOCOGTGGTATCGACAAG				
		TTATCTGCGATCGTGAATCGCGACCAAGACCTG		OGTORCTIONTCTRICCAG RECORDORATICORCIAGO				
	ACAGAGACTITC: G TCATOGECA: OGTCGACTCTGECAAGTCGACCACTGTGAGGACTCTCCTCGACAATGAGO JJ ACAGAGACTITC: G TCATOGECA: OGTCGACTCTGGCAAGTCGACCACTGTGAGGACTCTCCTCGACAATGAGO	ATATCTGCCATCGTCAATCCCGACCAAGACCTGC		IGGTCACTTGATCTACCAG TGCGGTGGTATCGACAAG				
		TTATCTGCCATCGTCAATCCCGACCAAGACCTGC ATATCTGCCATCGTCAATCCCGACCAAGACCTGC	T - GOOGTA TT TCTCAAAGTCAACA TACTGACATCG TT TCACAGAC C - GOOGTA TT TCTCAAAGTCAACA TACTGACATCG TT TCACAGAC	OGTCACTTGATCTACCAG RECORTOGTATCGATAAG				
1.2	ACAGAGACTITE & TEATEGRECA EGTEGACTETGREAAGTEGACCACTGTGAGTACTETECTEGACAATGAGO	ATATCTGCCATCGTCAATCCCGACCAAGACCTGC	C GOOGTA TETETEAAAGTEAACA TACTGACATOG TETEACAGACO	OGTCACTTGATCTACCAG TGCOGTGGTATCGATAAG				
		ATATCTGCCATCGTCAATCCCGACCAAGACCTG	C GGGGTA TT TCTCAAAGTCAACA TACTGACATOG TT TCACAGACO	IGGTCACTTIGATCTACCAG TGCGGTGGTATCGACAAG				
	AGAGAGACTITE: G TEATOGECA: OSTCOACTEGECAGISTEGACACTGIGAGIACTECTECTEGACAATGAGE 71 ACAGAGACTITE: G TEATOGECA: OSTCOACTEGECAGISTEGACTACTGIGAGIACTECTECTEGACAATGAGE	TTATCTGCCATCGTCAATCCCGACCAAGACCTG		DOTCACTEGATCTACCAG TOCODEGOTATCGATAAG				

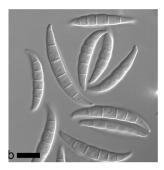
#### Takeaway ideas

- Oligotyping can extract additional information from your amplicons
  - Analyzes positional entropy in aligned reads
- It is first necessary to classify reads
  - dbcAmplicons
  - Qiime
  - Mothur
- Reads corresponding to a single species or genus can be aligned and oligotyped





### Overview



- 1. Introduction to my study system
  - 1. Fusarium
  - 2. Background on the translation elongation factor 1-alpha locus
- 2. Explanation of oligotyping: getting more from your amplicons
  - 1. Shannon's Entropy
  - 2. Oligotyping

#### 3. Pipeline & performance on mock communities

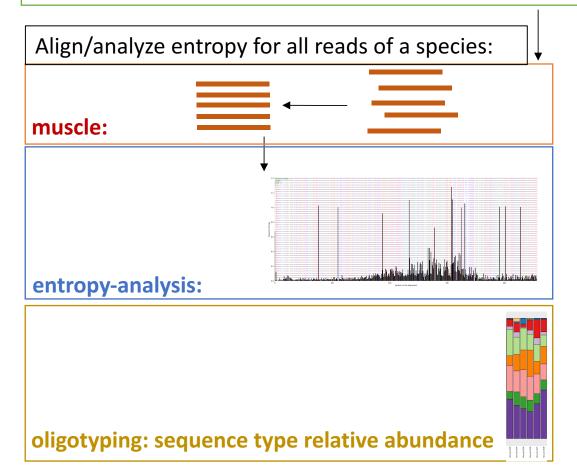
- 1. vsearch interspecific relative abundance
- 2. Oligotyping **intra**specific relative abundance
- 4. Where you can learn more

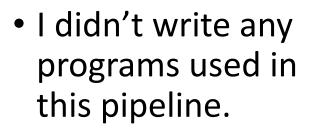
## Pipeline overview

- Bin sequences by barcode:
  - Join paired-end reads (pear)
  - Split reads by barcode (fastx\_toolkit)
  - Convert fastq to fasta format (fastx\_toolkit)
- Count reads for each species
  - Query samples with species-specific vsearch parameters
- Gather subspecies diversity
  - Extract sequences from a given species (e.g. *F. oxysporum*)
  - Align extracted sequences (muscle)
  - Trim primer/barcode regions manually (MEGA7)
  - Conduct entropy analysis (oligotype)
  - Generate oligotypes (oligotype)

Count reads for each *Fusarium* species: Sequences from sample A: Species queries: = 4 = 1 = 3

#### vsearch: obtains read counts per species by % identity







# VSEARCH: a versatile open source tool for metagenomics

Torbjørn Rognes<sup>1,2</sup>, Tomáš Flouri<sup>3,4</sup>, Ben Nichols<sup>5</sup>, Christopher Quince<sup>5,6</sup> and Frédéric Mahé<sup>7,8</sup>

**1792–1797** Nucleic Acids Research, 2004, Vol. 32, No. 5 DOI: 10.1093/nar/gkh340

MUSCLE: multiple sequence alignment with high accuracy and high throughput

Robert C. Edgar\*



British Ecological Society

Methods in Ecology and Evolution 2013, 4, 1111–1119

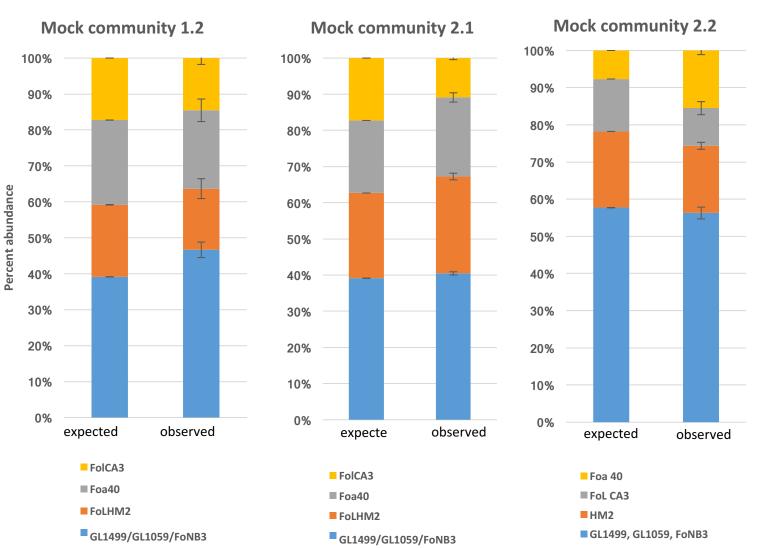
doi: 10.1111/2041-210X.12114

## Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data

A. Murat Eren\*, Loïs Maignien, Woo Jun Sul, Leslie G. Murphy, Sharon L. Grim, Hilary G. Morrison and Mitchell L. Sogin

## Estimating relative abundance: Oligotyping

- Mock communities:
  - 4 EF-1a sequence types
  - Fusarium oxysporum only
- Expected abundance calculated based on DNA concentration
- Averages taken from two technical reps
- Sequences are identical to the input



#### Where you can learn more

#### • Eren lab website:

- <u>http://merenlab.org/software/oligotyping/</u>
- Frontiers in Microbiology Research topic: New insights into microbial ecology through subtle nucleotide variation
  - <u>http://journal.frontiersin.org/researchtopic/2427</u> /new-insights-into-microbial-ecology-throughsubtle-nucleotide-variation



Methods in Ecology and Evolution 2013, 4, 1111–1119

ating between closely related

#### Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data

A. Murat Eren\*, Loïs Maignien, Woo Jun Sul, Leslie G. Murphy, Sharon L. Grim, Hilary G. Morrison and Mitchell L. Sogin The ISME Journal (2015) 9, 968–979 © 2015 International Society for Microbial Ecology All rights reserved 1751-7362/15 www.nature.com/ismei OPEN

**ORIGINAL ARTICLE** 

npg

C

doi: 10.1111/2041-210X.12114

Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of highthroughput marker gene sequences

A Murat Eren, Hilary G Morrison, Pamela J Lescault, Julie Reveillaud, Joseph H Vineis and Mitchell L Sogin Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA, USA