Enabling Cross-Species Computation on Phenotype in Plants

Carolyn J. Lawrence, Iowa State University

Preview

- Phenotype as a datatype
- Ontologies
- What's possible (examples)
- What we can do to predict plant biology
- Three slides
  - Current trends
  - Opportunities
  - Bottlenecks

Phenotype

- A **phenotype** (from Greek *phainein*, 'to show' + *typos*, 'type') is the composite of an organism's observable characteristics or traits, such as its morphology, development, biochemical or physiological properties, phenology, behavior, and products of behavior. A phenotype results from the expression of an organism's genes as well as the influence of environmental factors and the interactions between the two. -wikipedia

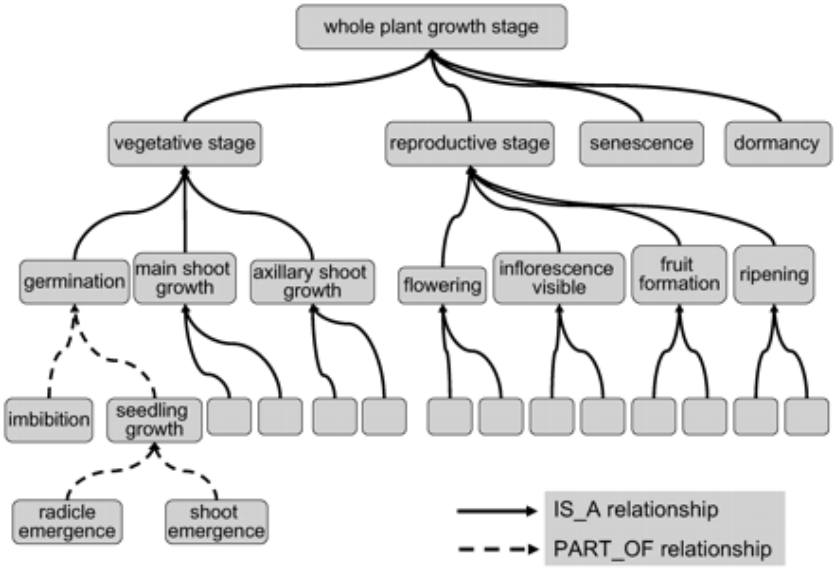
- Phenotype is EVERYTHING

Phenotype

- Extremely diverse data type (can range from expression profile to behavior)
- Associated to individuals, populations, or species
- Different levels (summary, measurement data)
- Can be comparative (mutant vs. wild type) or absolute (days to flowering of a cultivar)
- Data integration - needs extensive connections to other types of data (e.g., seed stocks, genes, experimental methods, publications)
- Database schema and interface design
- Data representation - how to represent the data in a consistent way across experiments, research communities, and species
- Data accessibility - how do we get data out of literature and into the database?

Ontologies

scholarsresource.com



Pujar et al., 2006 Plant Phys.

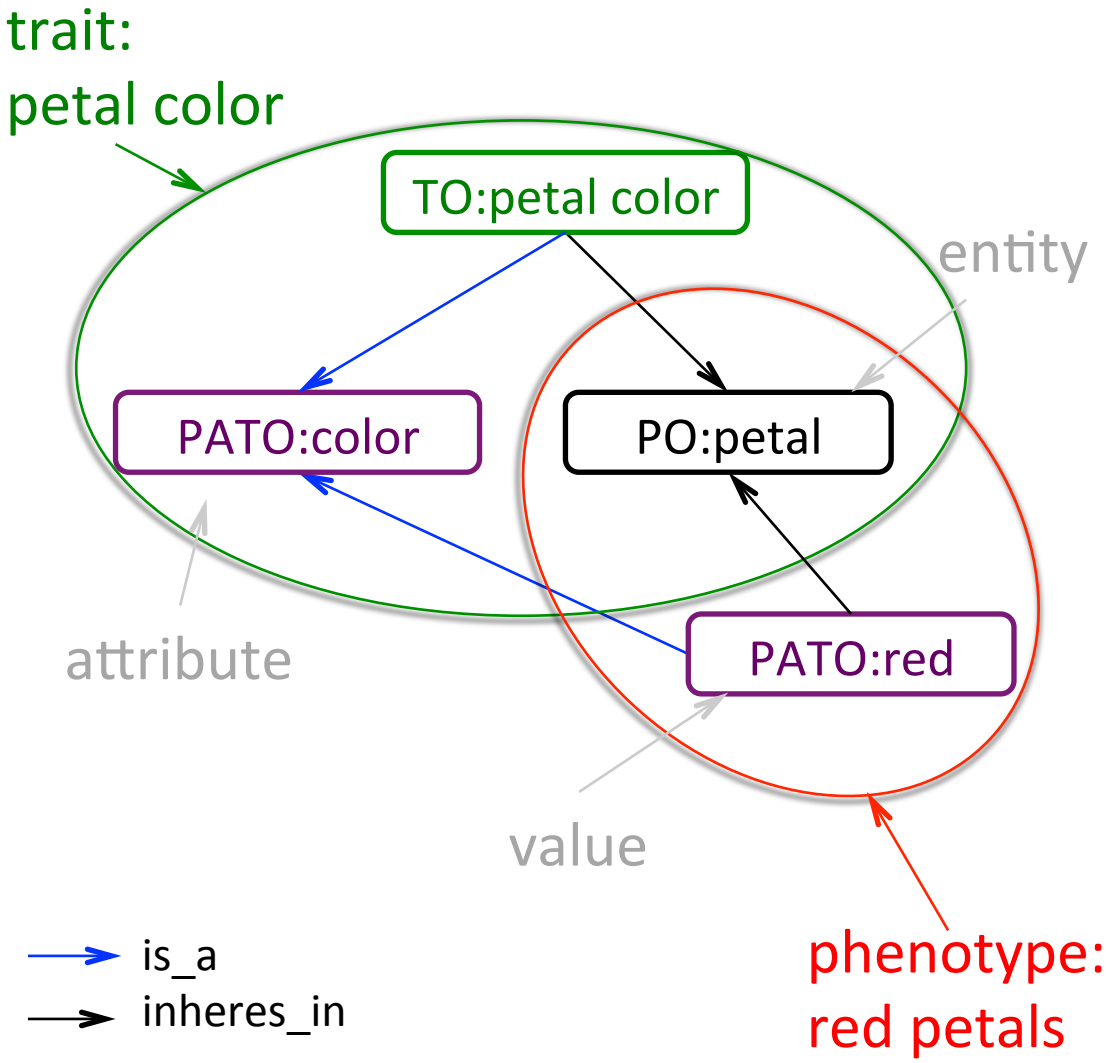
Metaphysics:
The study of being or existence

Biological applications: A structured vocabulary that includes definitions of terms in a domain and relationships among them

Traits versus phenotypes

A **trait** is a combination of an **entity** and an **attribute.**
Example: petal color

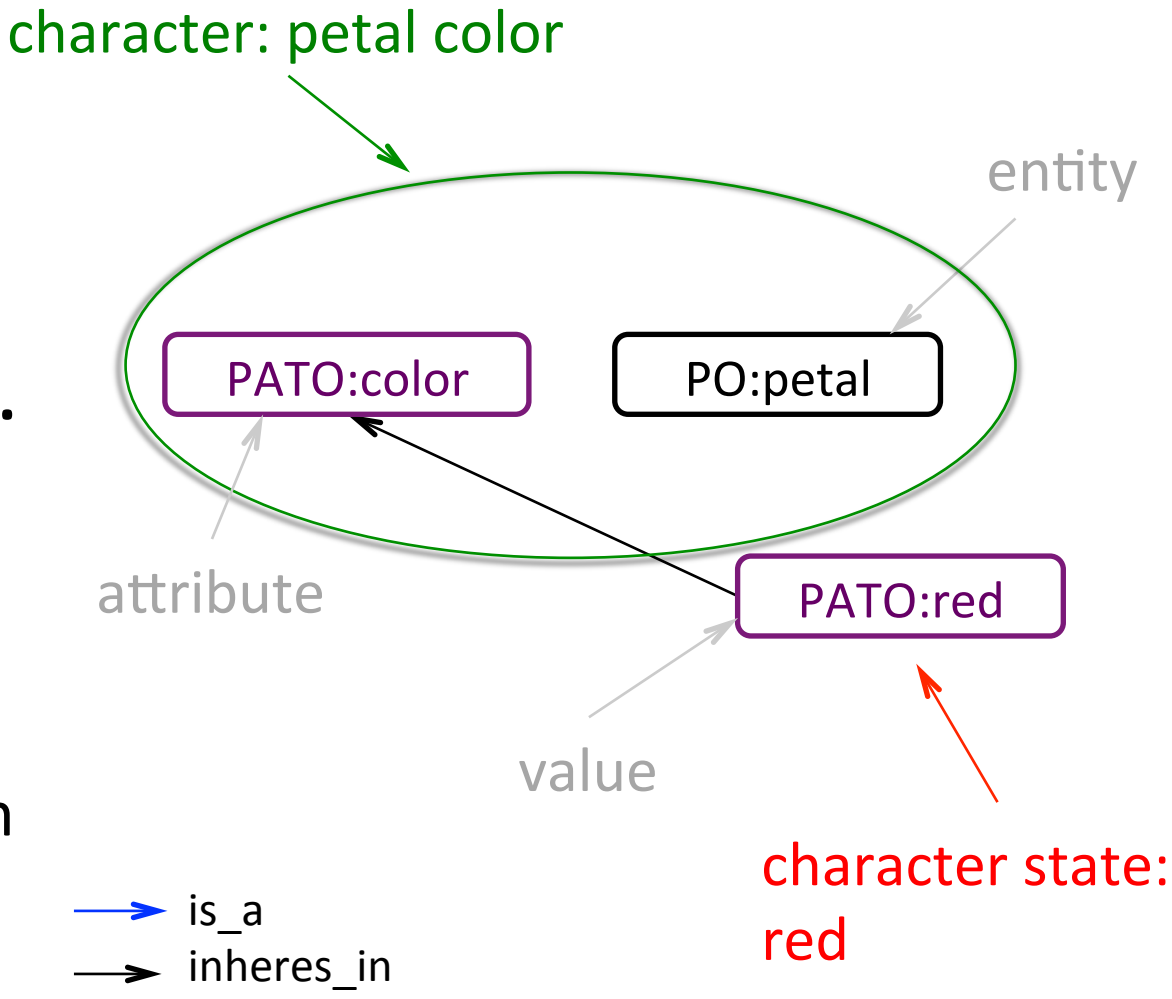
A **phenotype** is a combination of a **entity** and a **value**.
Example: red petals



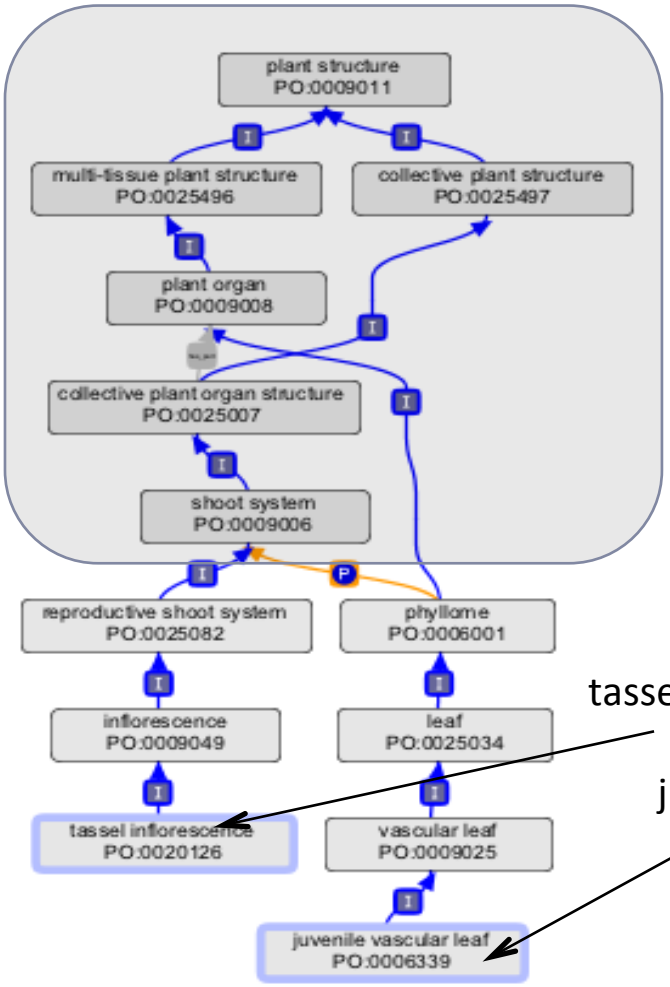
Characters and character states

A **character** is a combination of an **entity** and an **attribute.**
Example: petal color

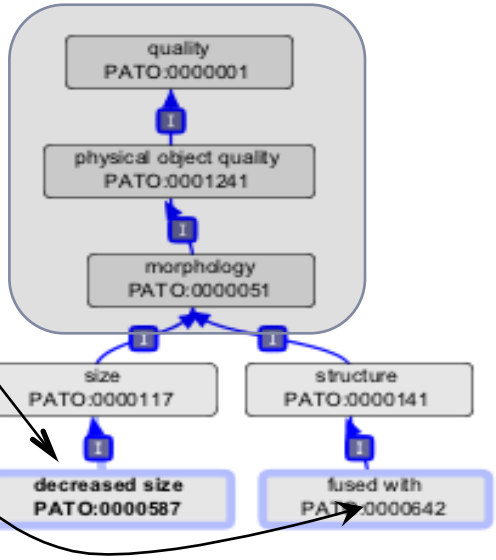
A **character state** is a **value**. Meaningless without the entity from the character
Example: red



What does this look like?



Semantic similarity of two phenotypes can be estimated based on the proportion of the ontology pathways that they share (grey boxes).



tassel inflorescence | decreased size

juvenile vascular leaf | fused

Entities Phenotypes Qualities

What does this look like?

Conversion to ontology statements:

Description of Mutant Phenotype*	Atomized Phenotype statements	Entity	Quality
Narrow leaves; Narrow, slightly elongated floral organs; Twisted siliques	Narrow leaves Narrow floral organs Elongated floral organs Twisted Siliques	PO: Vascular leaf PO: Floral organ PO: Floral organ PO: Fruit	PATO: Narrow PATO: Narrow PATO: Elongated PATO: Twisted
Delayed flowering; Reduction in total chlorophyll	Delayed flowering Reduction in total chlorophyll	GO: Flowering CHEBI: Chlorophyll	PATO: Delayed PATO: Decreased concentration

*From Lloyd and Meinke, Plant Physiol. 2012 Mar;158(3):1115-29

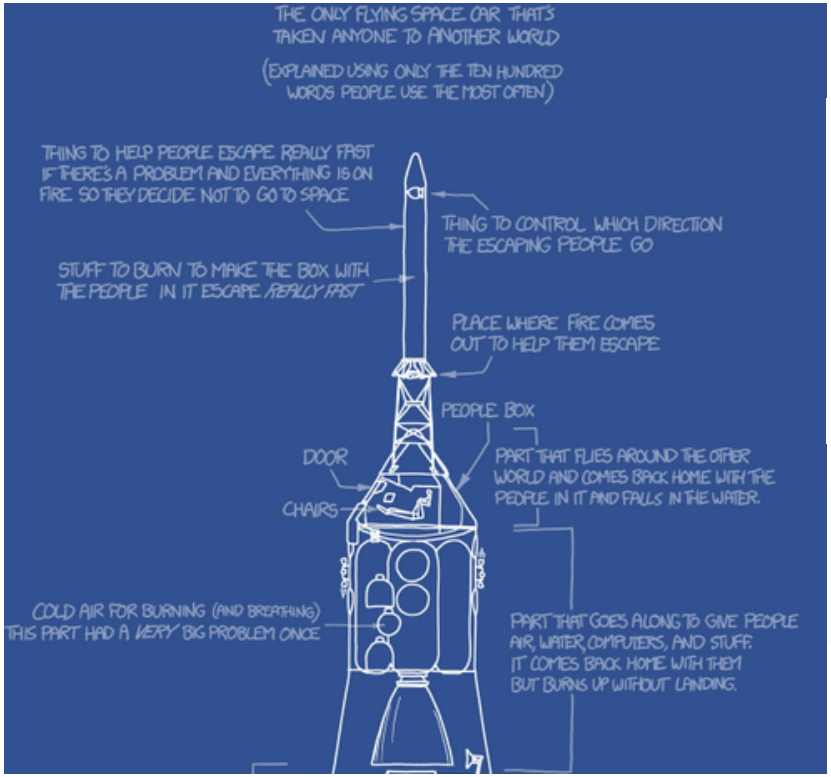
Does it matter what we call these entity/ quality associations?

- It doesn't really matter (to a computer) which word you use (trait, phenotype, or character), as long as you logically define your terms and specify the data structure

Aren't we losing something by using general terms?

THE UP-GOER FIVE TEXT EDITOR

CAN YOU EXPLAIN A HARD IDEA USING ONLY THE TEN HUNDRED MOST USED WORDS? IT'S NOT VERY EASY. TYPE IN THE BOX TO TRY IT OUT.



Carolyn J Lawrence shared a link.
September 27, 2013 near Ames, IA

http://xkcd.com/1133/ Can you explain your science with the ten-hundred most used words in the English language? Try it here: http://splasho.com/upgoer5/. I came up with this: "I help people to find out what others have learned about the most important green thing grown in the world."

James Schnable "I use computers to study how life turns light into power and how different types of life figured out how to do the same things." Not having 'plants' as a word is a major constraint.
September 27, 2013 at 8:47am · Edited · Unlike · 3

Why on EARTH would we want to go to all this trouble?!?

High-quality discovery potential!

- Identification, through the similarity of recorded phenotypes, of other alleles of the same gene, other members of a signaling pathway, and orthologous genes and pathway members across species: Washington et al. 2009, PLoS Biol 7(11): e1000247. -> Novel, testable predictions!
- Prediction of Arabidopsis negative gravitropism as a model for human Waardberg disease (ear development) from phenolog associations (ancient vesicle trafficking system): McGary et al. 2010, PNAS 107(14): 6544–6549. -> Identification of new, tractable model systems!
- By expansion to cross-species inference, predicting novel players in genetic networks is possible (e.g., response to vernalization) Woods et al. , BMC Bioinformatics 2013, 14:203. -> Functional prediction for uncharacterized genes!

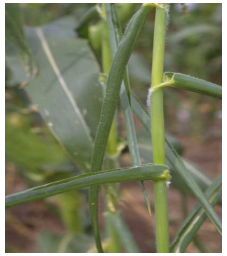
What we can do to predict plant biology

In the absence of sequence-based homology, shared multidimensional attributes (complex biologically relevant associations) among characterized and uncharacterized genes enables discovery of novel genes that potentially cause a given phenotype.

Dwarf plants




Rolled leaves




Plant Phenotype Pilot Project

Done or in progress:

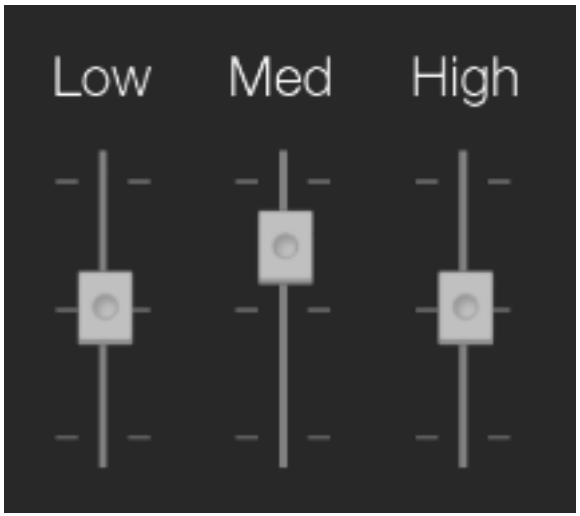
- Collect a broad set of phenotype descriptions for a representative set of plant species
- Identify ontologies to be used to describe phenotypes
- Develop standardized format and best practices to ensure consistency
- Translate phenotype descriptions into ontology statements

Next steps:

- Data analysis
  - Clustering of genes into pathways
  - Degree of correlation between sequence and phenotype
  - Computational prediction of gene candidates for uncloned mutant genes and QTL
- Apply lessons learned
  - Is the data set big enough?
  - Are the ontologies complete enough?
  - Is our annotation consistency good enough?
  - Better analysis methods?

Three slides:

- Current trends and state-of-the-art
- Opportunities
- Bottlenecks and needs



Current Trends and State of the Art

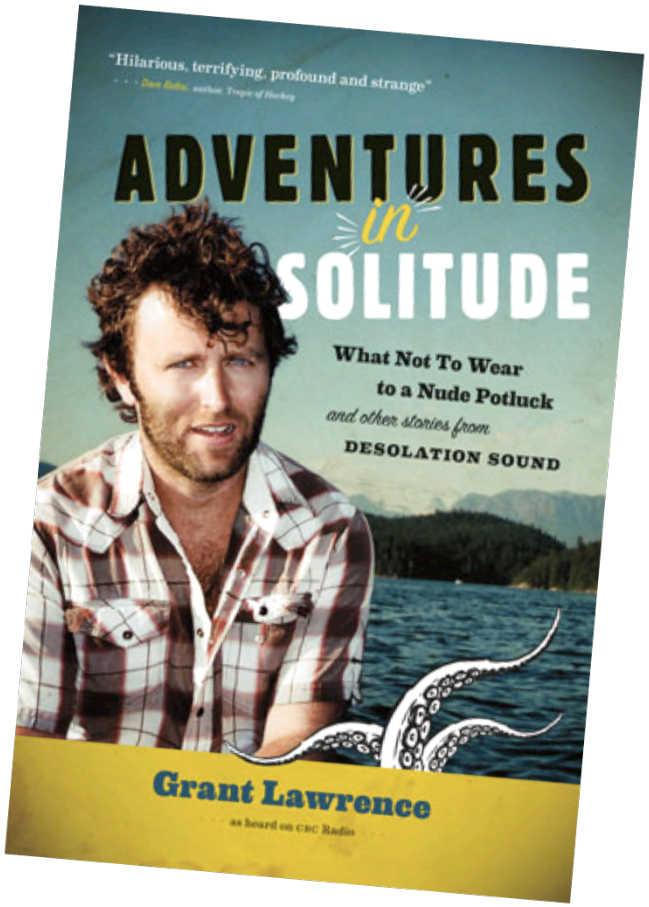
- High-throughput phenotyping within a species/crop
- Clear transition from thinking that genotype defines phenotype to recognizing phenotypic plasticity in varied environments and need to codify environment
- Interest and ability to compute on phenotype across domains of life and all described biological processes

Opportunities

- Discover and complete biological pathways and interaction networks
- Predict potential phenotypes for uncharacterized genes (guilt by association)
- Develop non-obvious models for biological processes across taxonomic boundaries

Bottlenecks and Needs

- Creating the dataset(s) and integrating them takes curators (AKA well-trained humans)
- Making data logically accessible takes domain experts working with computer scientists (AKA well-trained super-humans)
- Embracing complexity! 'Big Data' is not limited in scope to enormous file size…



Phenotypes are Big Data (?)

- **Big Data** is characterized as having extreme or variable values of one or more of the following characteristics:

Volume[1] (size)	Sequence, SNPs, expression data
Velocity[1] (acquisition rate)	Images, sequence
Variety[1] (structure)	Data formats, alternative standards
Variability[2] (in meaning)	Nomenclature, ontologies
Complexity[3] (in relationships)	Mutation to genotype to phenotype…
Veracity (quality or provenance)	Gold standard datasets
Volatility (changes over time)	Genome sequence assembly releases

1 Doug Laney, "3-D Data Management: Controlling Data Volume, Velocity, and Variety," 2001.
2 Brian Hopkins, "Blogging From the IBM Big Data Symposium - Big Is More Than Just Big," 2011.
3 Valentin T Sribar, et al., "'Big Data' Is Only the Beginning of Extreme Information Management," 2011.

Acknowledgements

Collaborators of the plant phenotype pilot project:

Ethalinda Cannon (USDA/ARS & MaizeGDB)
Steve Cannon (USDA/ARS & Soybase)
Laurel Cooper (Oregon State & Plant Ontology)
Jack Gardiner (Iowa State University)
George Gkoutos (University of Aberystwyth)
David Grant (USDA-ARS & Soybase)
Lisa Harper (USDA/ARS & MaizeGDB)
Mingze He (Iowa State University)
Eva Huala (Co-lead; Carnegie & TAIR)
Pankaj Jaiswal (Oregon State & Plant Ontology)
Scott Kalberer (USDA/ARS & Soybase)
Johnny Lloyd (Oklahoma State University)
David Meinke (Oklahoma State University)
Naama Menda (SGN & Cornell/BTI)
Lukas Mueller (SGN & Cornell/BTI)
Rex Nelson (USDA/ARS & Soybase)
Anika Oellrich (EBI)
Ramona Walls (iPlant)

Phenotype RCN

Phenotype Ontology Research Coordination Network

tair

maizeGDB
Maize Genetics and Genomics Database

SoyBase
Integrating Genetics and Molecular Biology for Soybean Researchers

sol genomics network

Plant Ontology PO Consortium

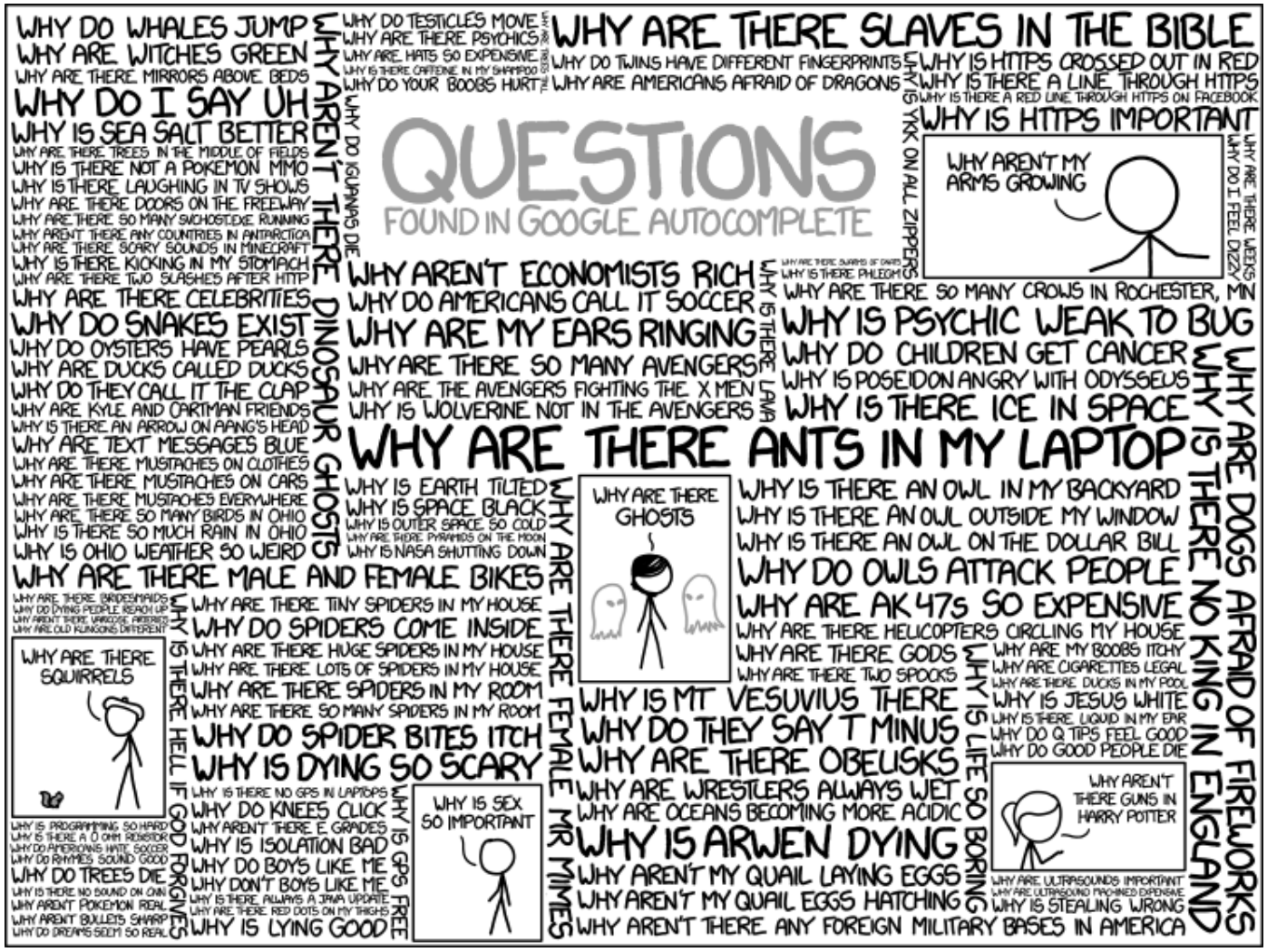
iPlant Collaborative™

NSF

USDA ARS

NCGA

Questions?


QUESTIONS
FOUND IN GOOGLE AUTOCOMPLETE
WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY ARE THERE MALE AND FEMALE BIKES
WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARICOSE AVERTS
WHY ARE OLD KLINGONS DIFFERENT
WHY ARE THERE SQUIRRELS
WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL
WHY DO TESTICLES MOVE
WHY ARE THERE PSYCHICS
WHY ARE HATS SO EXPENSIVE
WHY IS THERE CAFFEINE IN MY SHAMPOO
WHY DO YOUR BOOBS HURT
WHY ARE THERE SLAVES IN THE BIBLE
WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT
WHY AREN'T MY ARMS GROWING
WHY IS YKK ON ALL ZIPPERS
WHY DO I FEEL DIZZY
WHY ARE THERE WEEKS
WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS
WHY ARE THERE ANTS IN MY LAPTOP
WHY ARE THERE SWARMS OF GNATS
WHY IS THERE PHLEGM
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY IS PSYCHIC WEAK TO BUG
WHY DO CHILDREN GET CANCER
WHY IS POSEIDON ANGRY WITH ODYSSEUS
WHY IS THERE ICE IN SPACE
WHY IS THERE LAVA
WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN
WHY ARE THERE GHOSTS
WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS
WHY ARE MY BOOBS ITCHY
WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE
WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY
WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD
WHY IS SEX SO IMPORTANT
WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA
WHY AREN'T THERE GUNS IN HARRY POTTER
WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG
WHY DO IGUANAS DIE
WHY AREN'T THERE DINOSAUR GHOSTS
WHY ARE THERE FEMALE MR MIMES
WHY IS THERE HELL IF GOD FORGIVES
WHY IS GPS FREE
WHY IS LIFE SO BORING
WHY ARE THERE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND