Computational Biology

Phenotype Spaces and Graph Products

Marc Hellmuth

Department of Mathematics
Stockholm University

Biology (greek: bios ,live') Biology is a natural science concerned with the study of life and living organisms, including their structure, function, growth, evolution, distribution, identification and taxonomy.

In other words, want to understand the structure and processes in living organisms

A central point to understand living organisms is to understand their phenotype.

To describe a living organisms, we must be able to describe their phenotype.

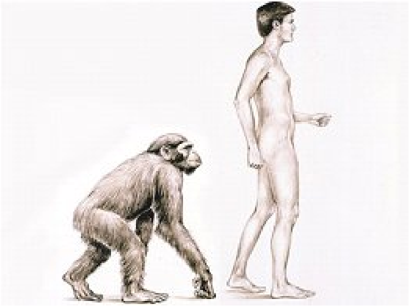
Biology (greek: bios ‚live') Biology is a natural science concerned with the study of life and living organisms, including their structure, function, growth, evolution, distribution, identification and taxonomy.

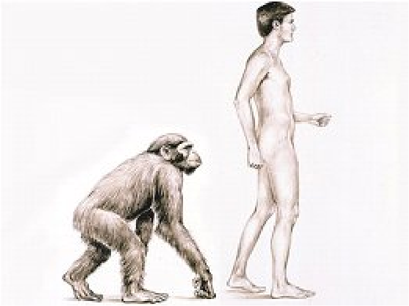
In other words, want to understand the structure and processes in living organisms

A central point to understand living organisms is to understand their phenotype.

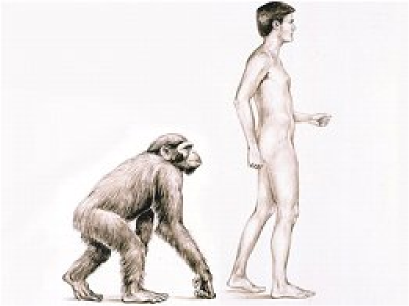
To describe a living organisms, we must be able to describe their phenotype.

Phenotype via Example: Great Ape (Menschenaffe) vs Human





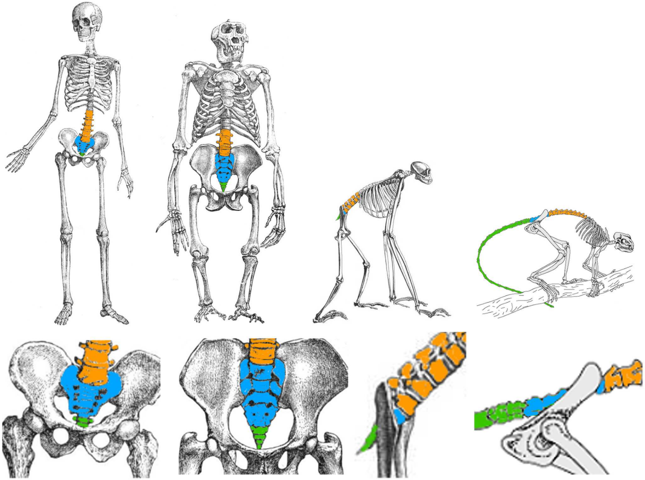
Many Similarities:
nose, eyes, 32 teeth, skeleton, bones, liver, mammal, . . .
But there are also big differences:
• hairiness
• apes don't have a nose bone
• Differences in skeleton → (Non)upright walking
• Humans have more effective spittle (= "water with enzyme") → improved predigestion → evolutionary benefit?



Phenotye is more that what one can see on a first glance. (skeleton, internal organs, production of enzymes, mammal, ..)

Phenotype = set of all traits (with its characteristics) of an organism.

What is a trait?



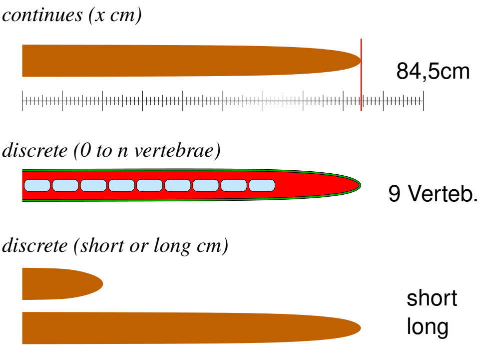
Wiki: All Apes (Hominoidea) have no tail. (.. mmh, tail or no tail?)
Hominoidea still have tails, but their characteristics differs e.g. from "Old World monkeys"

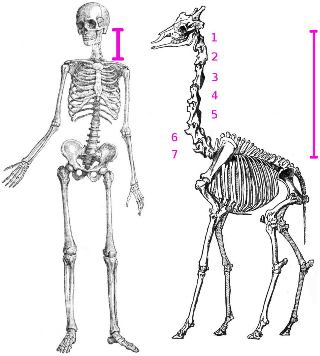
Phenotyope = set of all traits (with its characteristics) of an organism.

What is a trait?

- Traits can vary
- Traits are independent and can be freely combined
- Genes and their variants determine the characteristics of a trait

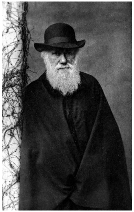
We consider the following “simplification”:

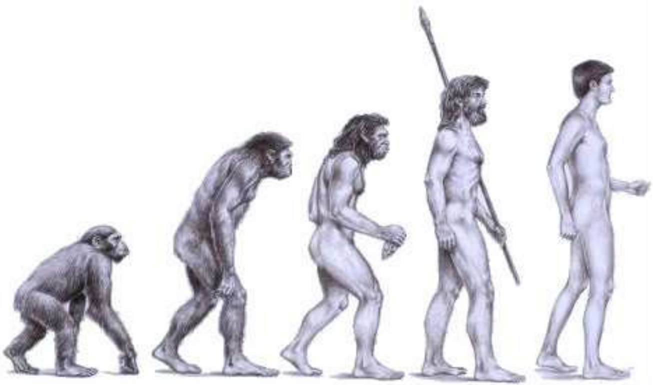




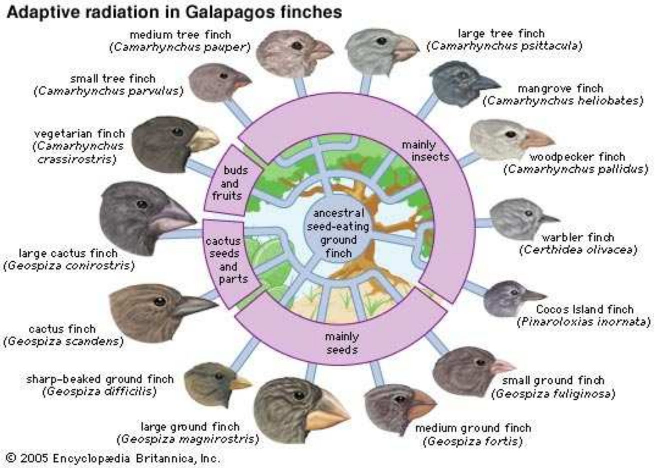
Judgening from the number of cervical vertebras the length of the neck of a giraffe and a human is equal.

"The Origin of Species" by Charles Darwin (1859)





About the variability of phenotypic characters.

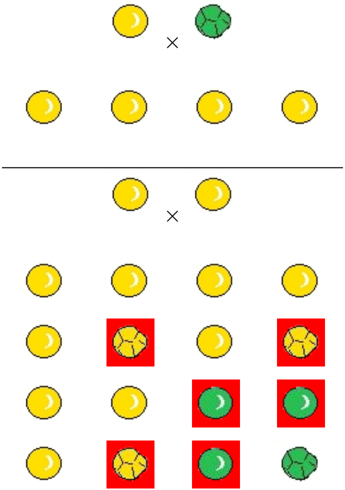


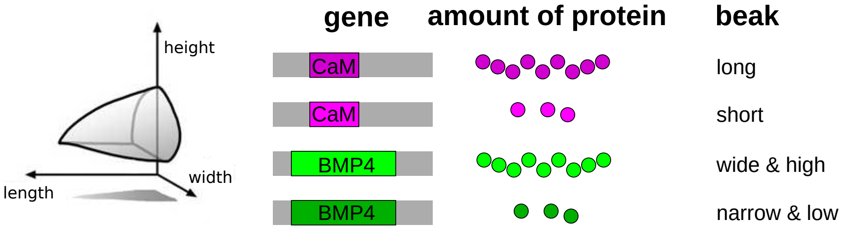
The shape of the beak is a character. Multiple characteristics exist.

Gregor Mendel (1822-1884) discovered that the characters

- **"Surface"**
  with the characteristics "smooth" and "wrinkled"
- and **"Seed color"**
  with the characteristics "yellow" and "green"

are independent and can be freely combined





Genes and their variants determine the characteristics of a character.

What is a character (Merkmal)?

Darwin's finches

Seeds Insects

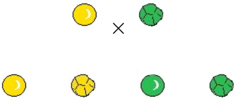
Characters can vary

Darwin's finches

Seeds Insects

Characters can vary

Mendel's peas



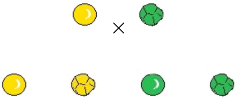
Characters are independent and can be freely combined

Darwin's finches

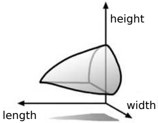
Seeds Insects

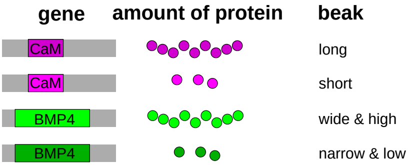
Characters can vary

Mendel's peas



Characters are independent and can be freely combined





Genes and their variants determine the characteristics of a character

Accessibility of genotypes based on "variational operators" $u \in \mathcal{U}$:

$$x \curvearrowleft_{\mathcal{U}} y$$

U-neighborhood of genotypes $y \in \mathbb{X}$:

$$N_{\mathcal{U}}(y) := \{x \in \mathbb{X} \mid x \curvearrowleft_{\mathcal{U}} y\}$$

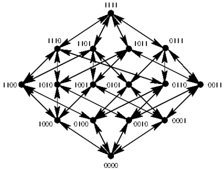
Genotype-Phenotype-Map
is mapping into phenotype space $\mathbb{P}$

$$f : \mathbb{X} \to \mathbb{P}$$

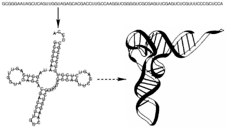
Accessibility of phenotypes $\alpha, \beta \in \mathbb{P}$:

$$\alpha \curvearrowleft_p \beta \iff \frac{|f^{-1}(\alpha) \cap N_{\mathcal{U}}(f^{-1}(\beta))|}{|N_{\mathcal{U}}(f^{-1}(\beta))|} > p$$

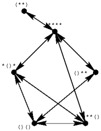
Genotype Space $\mathbb{X}$



GP-Map $f : \mathbb{X} \to \mathbb{P}$



Phenotype Space $\mathbb{P}$



$\mathcal{U} = \{\text{point-mutation}\}$ (just a quite simple example!)

$\mathbb{X} = \{AAT, GCT, GTT, CTT\}$ (genotypes)

$N_{\mathcal{U}}(y) := \{x \in \mathbb{X} \mid x \curvearrowleft_{\mathcal{U}} y\}$

- $N_{\mathcal{U}}(AAT) = \emptyset$
- $N_{\mathcal{U}}(GCT) = \{GTT\}$
- $N_{\mathcal{U}}(GTT) = \{GCT, CTT\}$
- $N_{\mathcal{U}}(CTT) = \{GCT\}$

Genotype-Phenotype-Map $f : \mathbb{X} \to \mathbb{P}$ with $\mathbb{P}$ = phenotype space (eg. sequence mapped to RNA-structure, protein, eye-color,...)

Accessibility of phenotypes $\alpha, \beta \in \mathbb{P}$:

$$\alpha \curvearrowleft_p \beta \iff \frac{|f^{-1}(\alpha) \cap N_{\mathcal{U}}(f^{-1}(\beta))|}{|N_{\mathcal{U}}(f^{-1}(\beta))|} > p$$

$\mathcal{U} = \{\text{point-mutation}\}$ (just a quite simple example!)

$\mathbb{X} = \{AAT, GCT, GTT, CTT\}$ (genotypes)

$N_{\mathcal{U}}(y) := \{x \in \mathbb{X} \mid x \curvearrowleft_{\mathcal{U}} y\}$

- $N_{\mathcal{U}}(AAT) = \emptyset$
- $N_{\mathcal{U}}(GCT) = \{GTT\}$
- $N_{\mathcal{U}}(GTT) = \{GCT, CTT\}$
- $N_{\mathcal{U}}(CTT) = \{GCT\}$

Genotype-Phenotype-Map $f : \mathbb{X} \to \mathbb{P}$ with $\mathbb{P}$ = phenotype space
(eg. sequence mapped to RNA-structure, protein, eye-color,...)

Accessibility of phenotypes $\alpha, \beta \in \mathbb{P}$:

$$\alpha \curvearrowleft_p \beta \iff \frac{|f^{-1}(\alpha) \cap N_{\mathcal{U}}(f^{-1}(\beta))|}{|N_{\mathcal{U}}(f^{-1}(\beta))|} > p$$

$\mathcal{U} = \{\text{point-mutation}\}$ (just a quite simple example!)

$\mathbb{X} = \{AAT, GCT, GTT, CTT\}$ (genotypes)

$N_{\mathcal{U}}(y) := \{x \in \mathbb{X} \mid x \curvearrowleft_{\mathcal{U}} y\}$

- $N_{\mathcal{U}}(AAT) = \emptyset$
- $N_{\mathcal{U}}(GCT) = \{GTT\}$
- $N_{\mathcal{U}}(GTT) = \{GCT, CTT\}$
- $N_{\mathcal{U}}(CTT) = \{GCT\}$

Genotype-Phenotype-Map $f : \mathbb{X} \to \mathbb{P}$ with $\mathbb{P}$ = phenotype space
(eg. sequence mapped to RNA-structure, protein, eye-color,...)

Accessibility of phenotypes $\alpha, \beta \in \mathbb{P}$:

$$\alpha \curvearrowleft_p \beta \iff \frac{|f^{-1}(\alpha) \cap N_{\mathcal{U}}(f^{-1}(\beta))|}{|N_{\mathcal{U}}(f^{-1}(\beta))|} > p$$

$\mathcal{U} = \{\text{point-mutation}\}$ (just a quite simple example!)

$\mathbb{X} = \{AAT, GCT, GTT, CTT\}$ (genotypes)

$N_{\mathcal{U}}(y) := \{x \in \mathbb{X} \mid x \curvearrowleft_{\mathcal{U}} y\}$

- $N_{\mathcal{U}}(AAT) = \emptyset$
- $N_{\mathcal{U}}(GCT) = \{GTT\}$
- $N_{\mathcal{U}}(GTT) = \{GCT, CTT\}$
- $N_{\mathcal{U}}(CTT) = \{GCT\}$

Genotype-Phenotype-Map $f : \mathbb{X} \to \mathbb{P}$ with $\mathbb{P}$ = phenotype space
(eg. sequence mapped to RNA-structure, protein, eye-color,...)

Accessibility of phenotypes $\alpha, \beta \in \mathbb{P}$:

$$\alpha \curvearrowleft_p \beta \iff \frac{|f^{-1}(\alpha) \cap N_{\mathcal{U}}(f^{-1}(\beta))|}{|N_{\mathcal{U}}(f^{-1}(\beta))|} > p$$

$\mathcal{U} = \{\text{point-mutation}\}$ (just a quite simple example!)

$\mathbb{X} = \{AAT, GCT, GTT, CTT\}$ (genotypes)

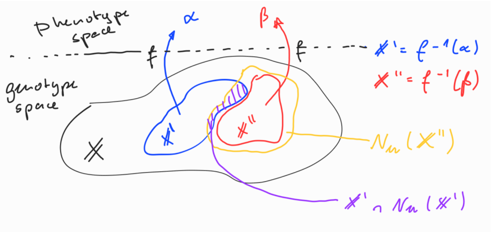
$N_{\mathcal{U}}(y) := \{x \in \mathbb{X} \mid x \curvearrowleft_{\mathcal{U}} y\}$

- $N_{\mathcal{U}}(AAT) = \emptyset$
- $N_{\mathcal{U}}(GCT) = \{GTT\}$
- $N_{\mathcal{U}}(GTT) = \{GCT, CTT\}$
- $N_{\mathcal{U}}(CTT) = \{GCT\}$

Genotype-Phenotype-Map $f : \mathbb{X} \to \mathbb{P}$ with $\mathbb{P}$ = phenotype space
(eg. sequence mapped to RNA-structure, protein, eye-color,. . .)

Accessibility of phenotypes $\alpha, \beta \in \mathbb{P}$:

$$\alpha \curvearrowleft_p \beta \iff \frac{|f^{-1}(\alpha) \cap N_{\mathcal{U}}(f^{-1}(\beta))|}{|N_{\mathcal{U}}(f^{-1}(\beta))|} > p$$



$\mathcal{U} = \{\text{point-mutation}\}$ (just a quite simple example!)

$\mathbb{X} = \{AAT, GCT, GTT, CTT\}$ (genotypes)

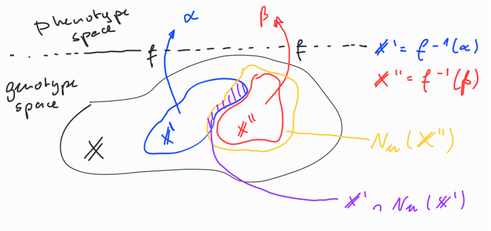
$N_{\mathcal{U}}(y) := \{x \in \mathbb{X} \mid x \curvearrowleft_{\mathcal{U}} y\}$

- $N_{\mathcal{U}}(AAT) = \emptyset$
- $N_{\mathcal{U}}(GCT) = \{GTT\}$
- $N_{\mathcal{U}}(GTT) = \{GCT, CTT\}$
- $N_{\mathcal{U}}(CTT) = \{GCT\}$

Genotype-Phenotype-Map $f : \mathbb{X} \to \mathbb{P}$ with $\mathbb{P}$ = phenotype space
(eg. sequence mapped to RNA-structure, protein, eye-color,. . .)

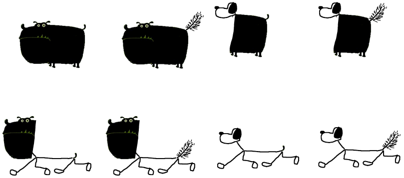
Accessibility of phenotypes $\alpha, \beta \in \mathbb{P}$:

$$\alpha \curvearrowleft_p \beta \iff \frac{|f^{-1}(\alpha) \cap N_{\mathcal{U}}(f^{-1}(\beta))|}{|N_{\mathcal{U}}(f^{-1}(\beta))|} > p$$



Given:

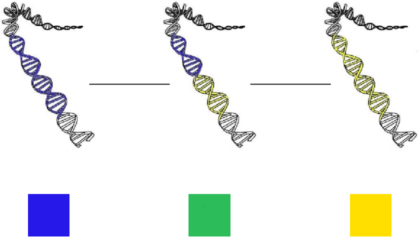
- Genomic Sequences from individual Organisms
- Genotyp-Phenotype-Map

Observable Phenotype	Possible Traits
	structure, color, sugar content, . . .
	tail, body, head, . . .

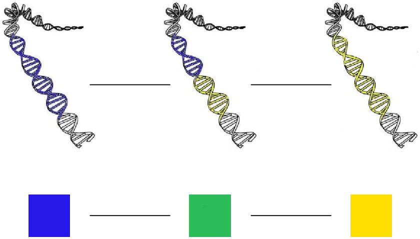
- Are putative traits independent?
- How can one determine traits?

$\longrightarrow$ Graphproducts

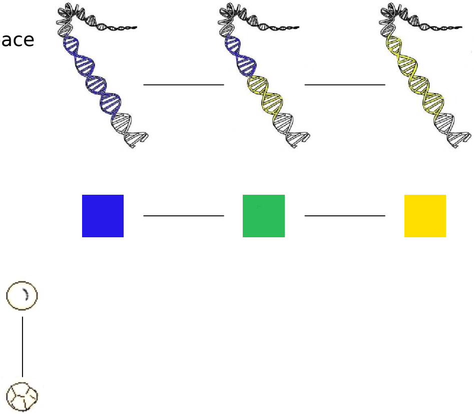
Genotype Space

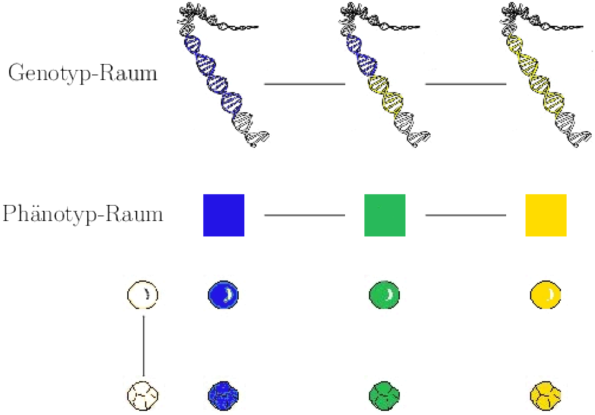


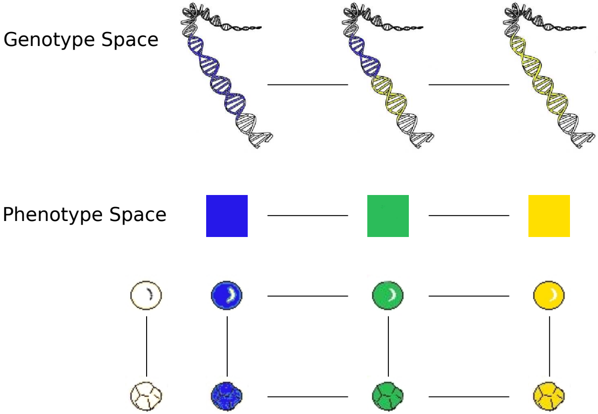
Genotype Space



Genotype Space

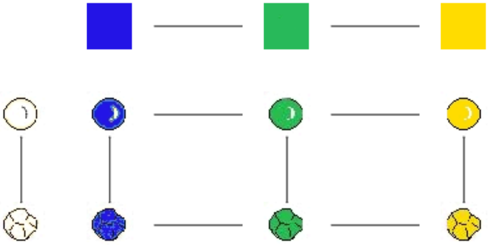



Genotyp-Raum
Phänotyp-Raum


Genotype Space
Phenotype Space

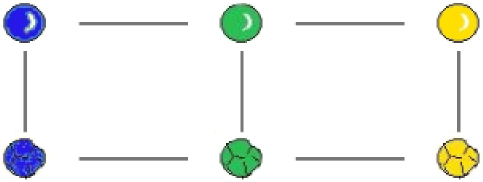
Theorem

Characters can vary independently $\iff$
They correspond to (local) prime factors of phenotype space



Quasi-Independence, Homology and the Unity of Type: A Topological Theory of Characters, Günter Wagner and Peter F. Stadler, J. theor. Biol., 2003

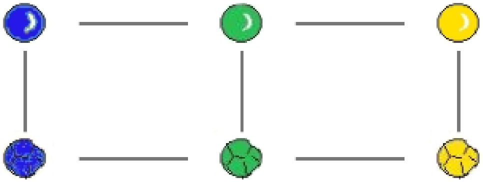
Problem: We may have a phenotype space, but we don't know the prime factors!



In the following, we will

- Define one (of the many possible) graph product: Cartesian product
- Define the term prime factors
- Sketch how to find prime factors

Problem: We may have a phenotype space, but we don't know the prime factors!



In the following, we will

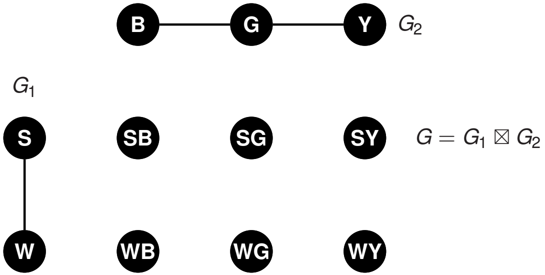
- Define one (of the many possible) graph product: Cartesian product
- Define the term prime factors
- Sketch how to find prime factors

As numbers, one can multiply graphs.

Given $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$:

$$V(G_1 \square G_2) = \{(v_1, v_2) \mid v_1 \in V(G_1), v_2 \in V(G_2)\}$$

.



commutative; associative; unique unit K_1: $G \square K_1 \simeq K_1 \square G \simeq G$

As numbers, one can multiply graphs.

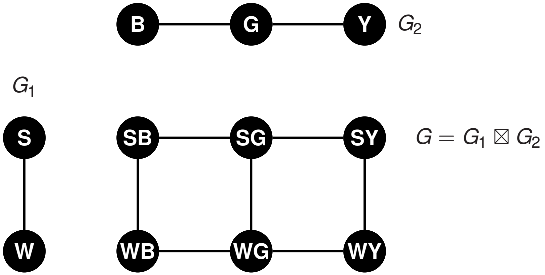
Given $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$:

$$V(G_1 \square G_2) = \{(v_1, v_2) \mid v_1 \in V(G_1), v_2 \in V(G_2)\}$$

.

Two vertices (x_1, x_2), (y_1, y_2) in $G_1 \square G_2$ are linked by an edge if:

$\{x_1, y_1\} \in E(G_1)$ and $x_2 = y_2$ or $\{x_2, y_2\} \in E(G_2)$ and $x_1 = y_1$



commutative; associative; unique unit K_1: $G \square K_1 \simeq K_1 \square G \simeq G$

As numbers, one can multiply graphs.

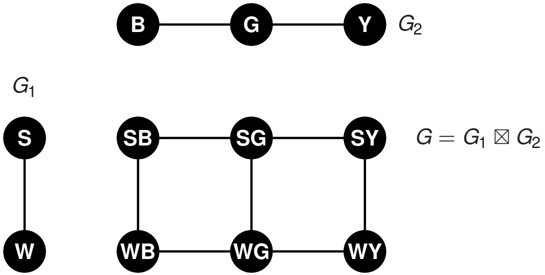
Given $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$:

$$V(G_1 \square G_2) = \{(v_1, v_2) \mid v_1 \in V(G_1), v_2 \in V(G_2)\}$$

.

Two vertices $(x_1, x_2), (y_1, y_2)$ in $G_1 \square G_2$ are linked by an edge if:

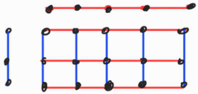
$\{x_1, y_1\} \in E(G_1)$ and $x_2 = y_2$ or $\{x_2, y_2\} \in E(G_2)$ and $x_1 = y_1$

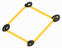


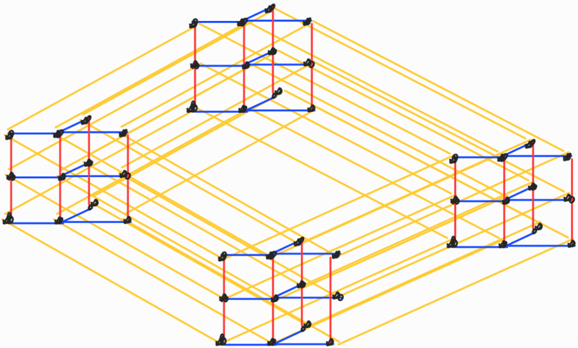
commutative; associative; unique unit K_1: $G \square K_1 \simeq K_1 \square G \simeq G$

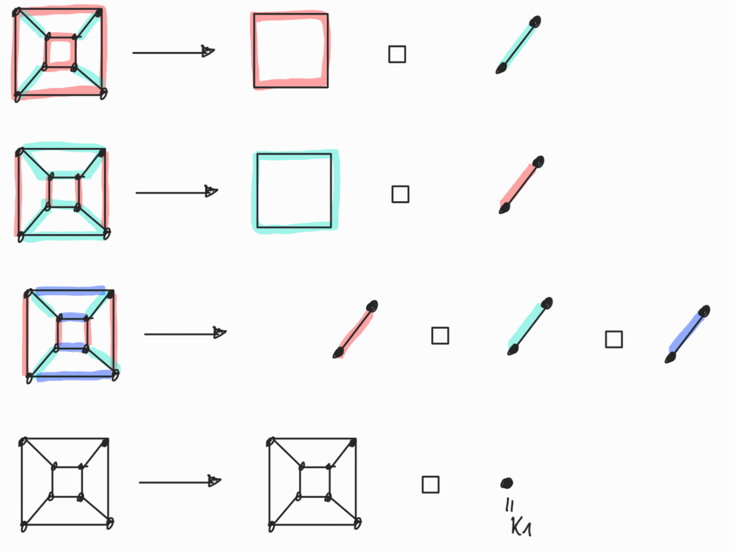
$\square_{i=1}^{n} K_2 = Q_n$: $Q_1 =$ , $Q_2 =$ , $Q_3 =$ , ...

$\text{path}_n \times \text{path}_m = n \times m$ grid :

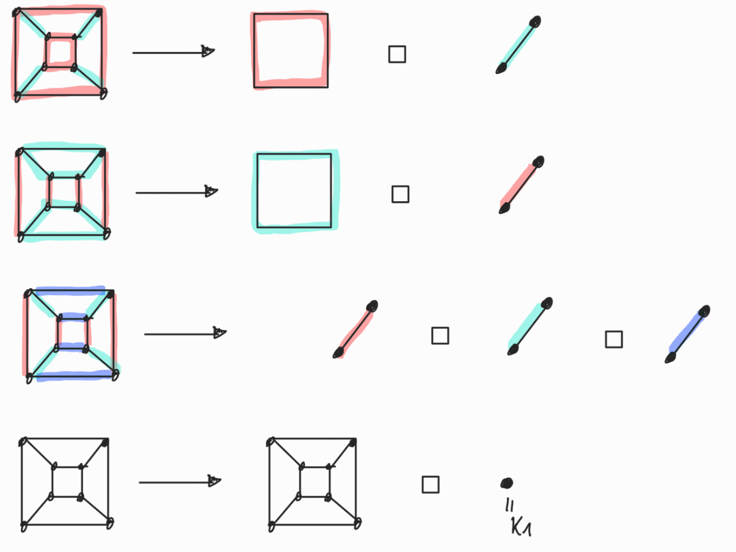


 $\square$  $\square$  :





Different factorization are possible and we are interested in the “finest one” = prime factorization



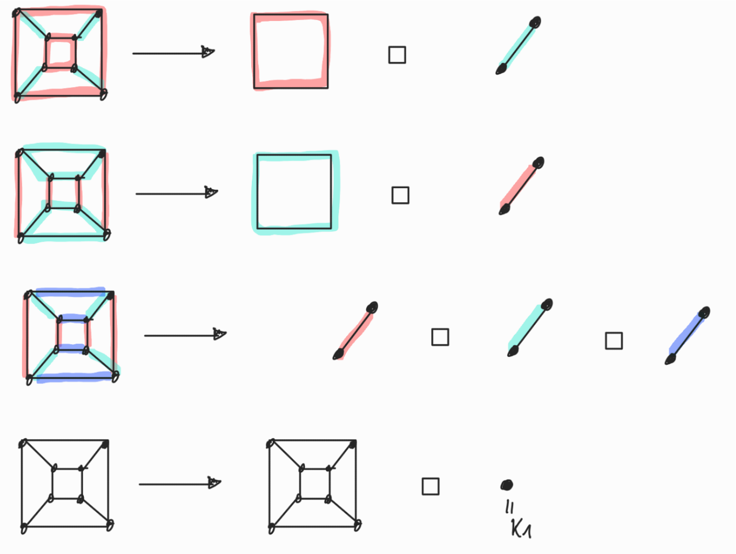
Different factorization are possible and we are interested in the "finest one" = prime factorization

G is prime, if

$$G = G_1 \boxtimes G_2 \quad \Rightarrow \quad G_1 \simeq K_1 \text{ or } G_2 \simeq K_1$$

PFD w.r.t. $\Box$ is unique.

Aim: Find PFD of given graphs G.



Essentially, we want to find a coloring $c\colon E \to C$ of the edges in the product that gives the copies of the prime factors.

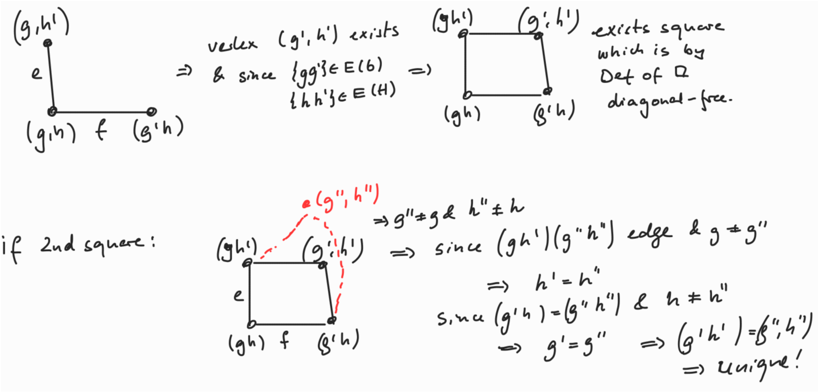
Lemma (Square-Property)

If e and f are incident edges in $G\Box H$ that are located in copies of different factors, then there is a unique diagonal-free square in $G\Box H$ that containes both e and f

Lemma (Square-Property)

If e and f are incident edges in G□H that are located in copies of different factors, then there is a unique diagonal-free square in G□H that containes both e and f

PROOF:



Lemma (Square-Property)

If e and f are incident edges in $G\Box H$ that are located in copies of different factors, then there is a unique diagonal-free square in $G\Box H$ that containes both e and f

By contraposition, if e and f are incident and there is **no** or **more than one** square that contains e and f, they must be in the copy of one and the same factor.

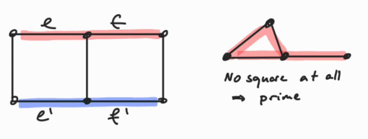
$\implies$ in this case, color $c(e) = c(f)$

Lemma (Square-Property)

If e and f are incident edges in G□H that are located in copies of different factors, then there is a unique diagonal-free square in G□H that containes both e and f

By contraposition, if e and f are incident and there is **no** or **more than one** square that contains e and f, they must be in the copy of one and the same factor.

$\Longrightarrow$ in this case, color $c(e) = c(f)$

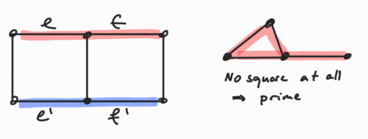


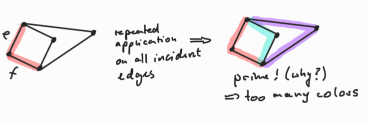
Lemma (Square-Property)

If e and f are incident edges in G□H that are located in copies of different factors, then there is a unique diagonal-free square in G□H that containes both e and f

By contraposition, if e and f are incident and there is **no** or **more than one** square that contains e and f, they must be in the copy of one and the same factor.

$\implies$ in this case, color $c(e) = c(f)$



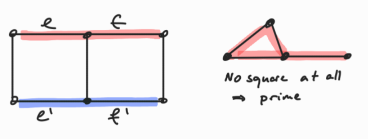


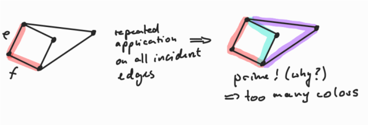
Lemma (Square-Property)

If e and f are incident edges in G□H that are located in copies of different factors, then there is a unique diagonal-free square in G□H that containes both e and f

By contraposition, if e and f are incident and there is **no** or **more than one** square that contains e and f, they must be in the copy of one and the same factor.

$\implies$ in this case, color $c(e) = c(f)$





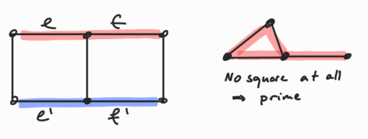
If e and f are opposite edges of a diag-free square, they belong to copy of one and the same factor.

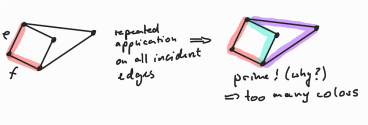
Lemma (Square-Property)

If e and f are incident edges in G□H that are located in copies of different factors, then there is a unique diagonal-free square in G□H that containes both e and f

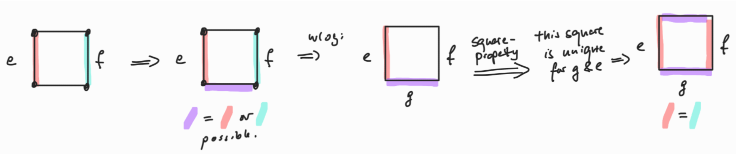
By contraposition, if e and f are incident and there is **no** or **more than one** square that contains e and f, they must be in the copy of one and the same factor.

$\Longrightarrow$ in this case, color $c(e) = c(f)$





If e and f are opposite edges of a diag-free square, they belong to copy of one and the same factor.



Basic Idea for computing PFD:

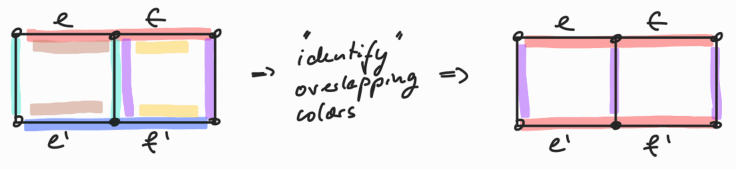
Put $c(e) = c(f)$ for all edges e, f that satisfy

- e and f are opposite edges of a square
- e and f are incident but do not lie in a common square

Basic Idea for computing PFD:

Put $c(e) = c(f)$ for all edges e, f that satisfy

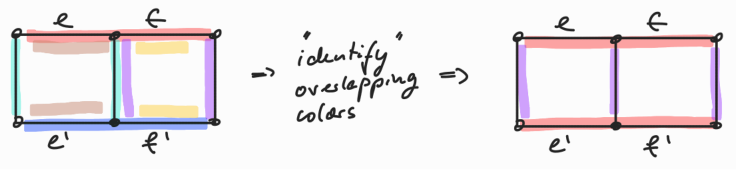
- e and f are opposite edges of a square
- e and f are incident but do not lie in a common square



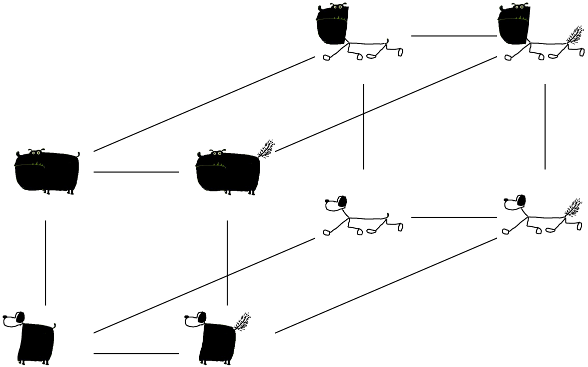
Basic Idea for computing PFD:

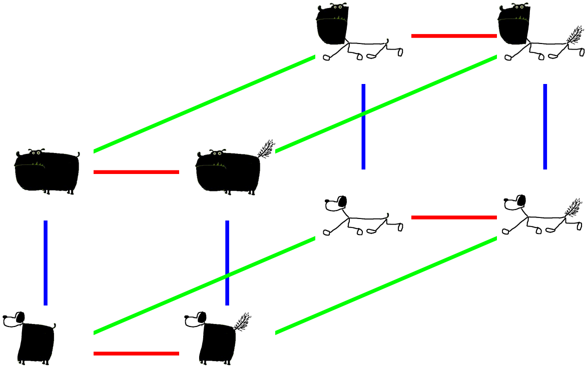
Put $c(e) = c(f)$ for all edges e, f that satisfy

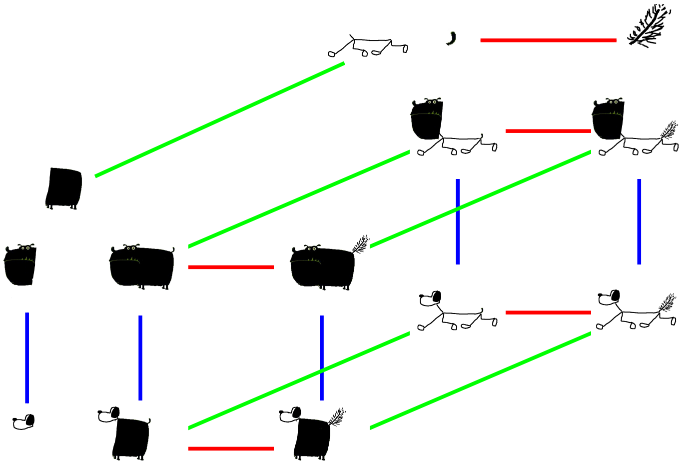
- e and f are opposite edges of a square
- e and f are incident but do not lie in a common square

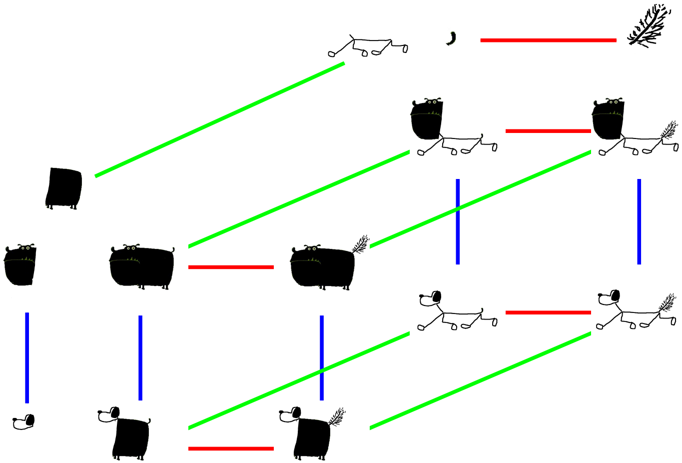


PFD can be computed in linear time.









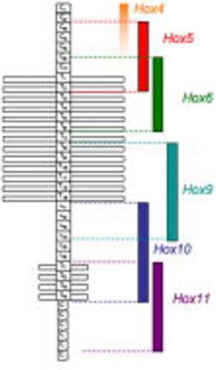
Comic VS Reality: tail = trait?

HOX genes determine the number and type of vertebras.

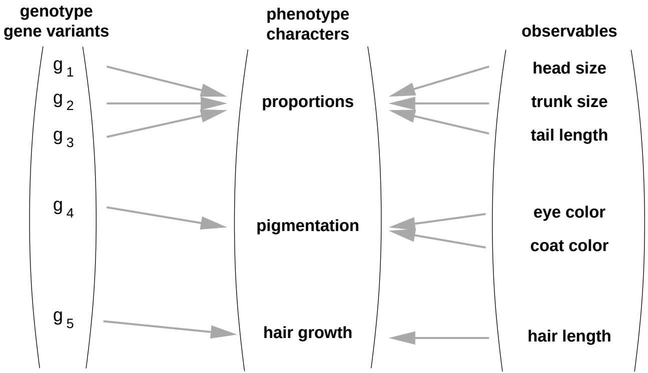
HOX genes are responsible for axial patterning of

- spinal column
- extremities
- brain and spinal cord
- uterus

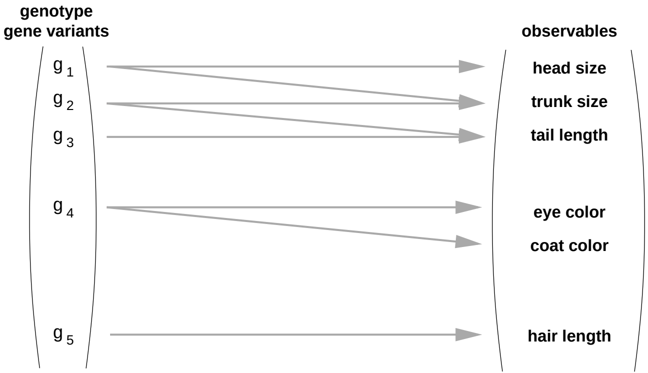
The length of the tail is, therefore, not an independent character.



Genes control the length of the tail



Head size, trunk size, and tail length are dependent.



What you see is what you abstract!

Genotype + Environment ⇔ Phenotype

Carotinoides in a flamingo's diet color its feathering.

Genes alone do not determine everything!