Utilization of knowledge graphs to consolidate biological context for non-model plants

Presenter: Brent Murphy

BASF
We create chemistry

Agricultural Solutions is an integral part of BASF Group

Core businesses

Chemicals

- Petrochemicals
- Intermediates

Materials

- Performance Materials
- Monomers

Industrial Solutions

- Dispersions & Resins
- Performance Chemicals

Nutrition & Care

- Care Chemicals
- Nutrition & Health

2025
BASF sales
€59.7 billion
Global R&D expenditures
€2.0 billion

Standalone businesses

Surface Technologies

- Environmental Catalyst and Metal Solutions

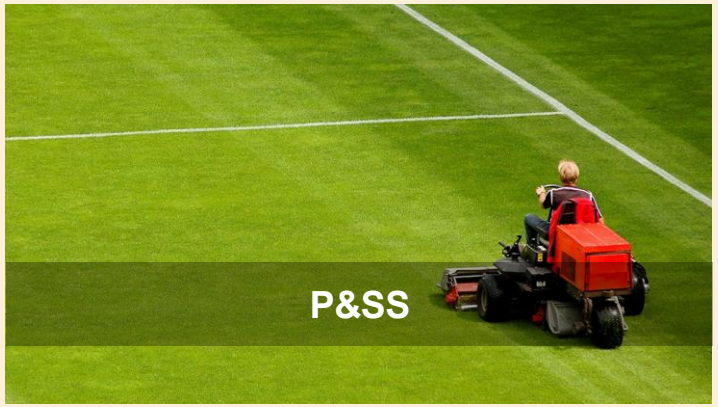
- Battery Materials

- Coatings

Agricultural Solutions

2025
AP sales
€9.6 billion
R&D expenditures
€1.0 billion

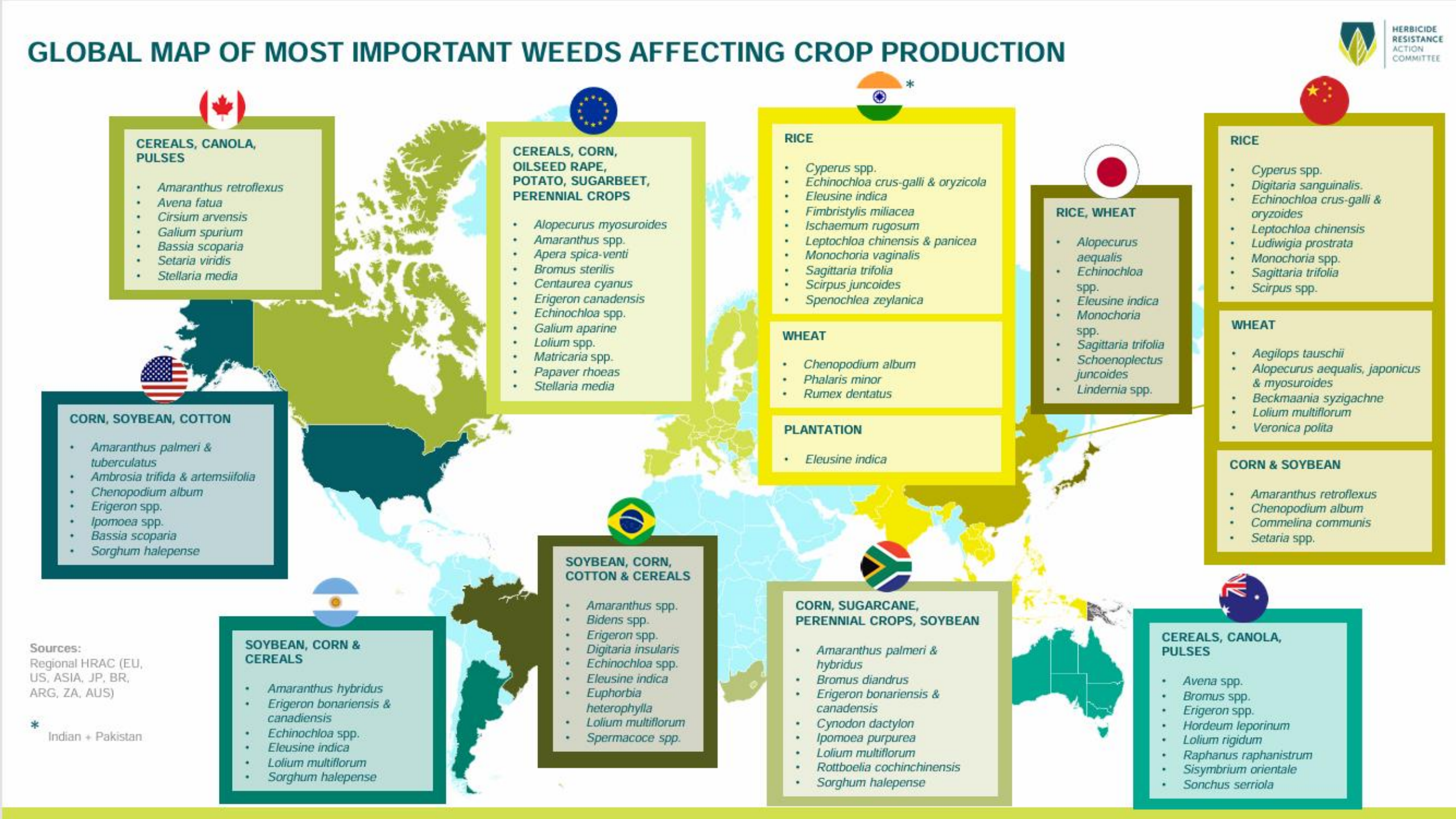
Our AP portfolio is continuously developing



BASF
We create chemistry

GLOBAL MAP OF MOST IMPORTANT WEEDS AFFECTING CROP PRODUCTION

HERBICIDE RESISTANCE ACTION COMMITTEE



CEREALS, CANOLA, PULSES

- *Amaranthus retroflexus*
- *Avena fatua*
- *Cirsium arvensis*
- *Galium spurium*
- *Bassia scoparia*
- *Setaria viridis*
- *Stellaria media*

CEREALS, CORN, OILSEED RAPE, POTATO, SUGARBEET, PERENNIAL CROPS

- *Alopecurus myosuroides*
- *Amaranthus* spp.
- *Apera spica-venti*
- *Bromus sterilis*
- *Centaurea cyanus*
- *Erigeron canadensis*
- *Echinochloa* spp.
- *Galium aparine*
- *Lolium* spp.
- *Matricaria* spp.
- *Papaver rhoeas*
- *Stellaria media*

*

RICE

- *Cyperus* spp.
- *Echinochloa crus-galli & oryzicola*
- *Eleusine indica*
- *Fimbristylis miliacea*
- *Ischaemum rugosum*
- *Leptochloa chinensis & panicea*
- *Monochoria vaginalis*
- *Sagittaria trifolia*
- *Scirpus juncoides*
- *Spenochlea zeylanica*

WHEAT

- *Chenopodium album*
- *Phalaris minor*
- *Rumex dentatus*

PLANTATION

- *Eleusine indica*

RICE, WHEAT

- *Alopecurus aequalis*
- *Echinochloa* spp.
- *Eleusine indica*
- *Monochoria* spp.
- *Sagittaria trifolia*
- *Schoenoplectus juncoides*
- *Lindernia* spp.

RICE

- *Cyperus* spp.
- *Digitaria sanguinalis.*
- *Echinochloa crus-galli & oryzoides*
- *Leptochloa chinensis*
- *Ludiwigia prostrata*
- *Monochoria* spp.
- *Sagittaria trifolia*
- *Scirpus* spp.

WHEAT

- *Aegilops tauschii*
- *Alopecurus aequalis, japonicus & myosuroides*
- *Beckmaania syzigachne*
- *Lolium multiflorum*
- *Veronica polita*

CORN & SOYBEAN

- *Amaranthus retroflexus*
- *Chenopodium album*
- *Commelina communis*
- *Setaria* spp.

CORN, SOYBEAN, COTTON

- *Amaranthus palmeri & tuberculatus*
- *Ambrosia trifida & artemsiifolia*
- *Chenopodium album*
- *Erigeron* spp.
- *Ipomoea* spp.
- *Bassia scoparia*
- *Sorghum halepense*

SOYBEAN, CORN & CEREALS

- *Amaranthus hybridus*
- *Erigeron bonariensis & canadiensis*
- *Echinochloa* spp.
- *Eleusine indica*
- *Lolium multiflorum*
- *Sorghum halepense*

SOYBEAN, CORN, COTTON & CEREALS

- *Amaranthus* spp.
- *Bidens* spp.
- *Erigeron* spp.
- *Digitaria insularis*
- *Echinochloa* spp.
- *Eleusine indica*
- *Euphorbia heterophylla*
- *Lolium multiflorum*
- *Spermacoce spp.*

CORN, SUGARCANE, PERENNIAL CROPS, SOYBEAN

- *Amaranthus palmeri & hybridus*
- *Bromus diandrus*
- *Erigeron bonariensis & canadensis*
- *Cynodon dactylon*
- *Ipomoea purpurea*
- *Lolium multiflorum*
- *Rottboelia cochinchinensis*
- *Sorghum halepense*

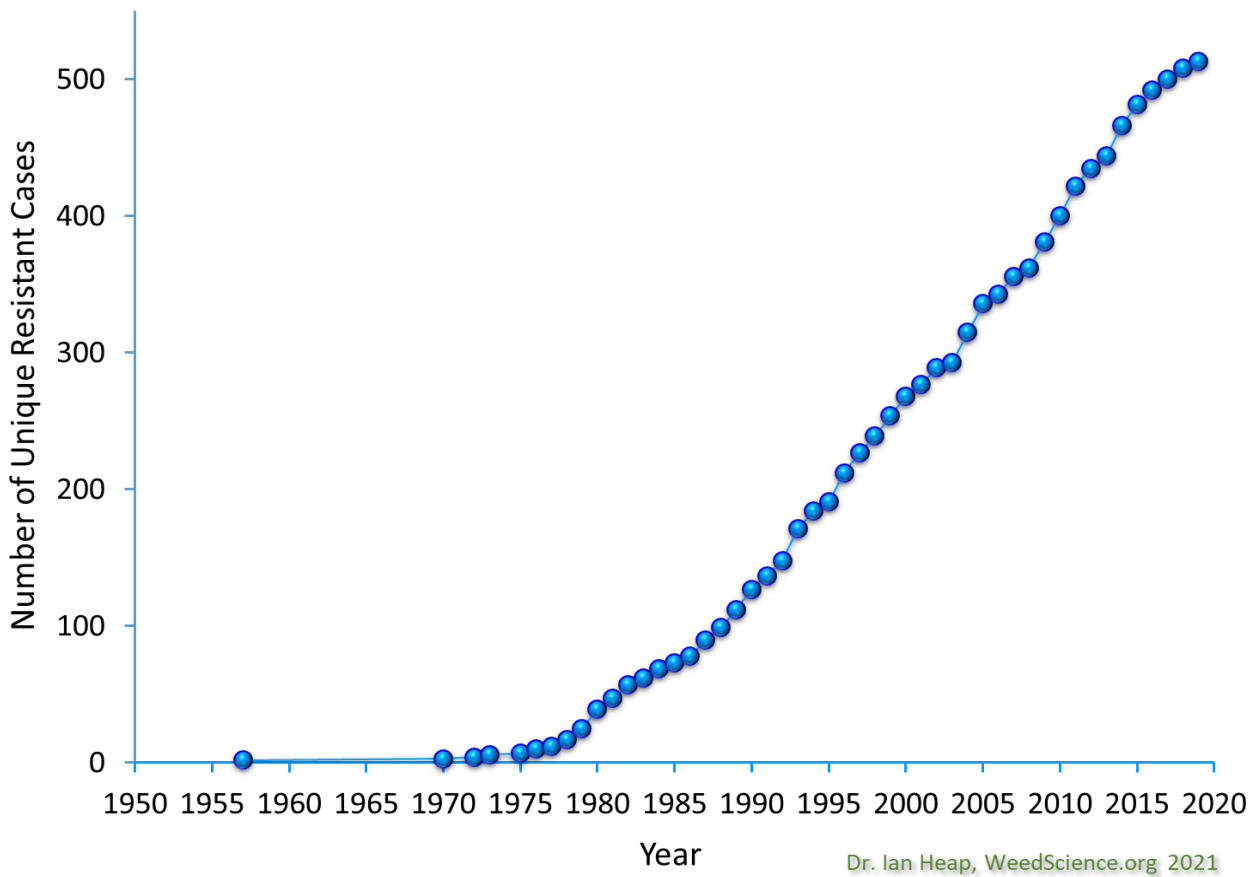
CEREALS, CANOLA, PULSES

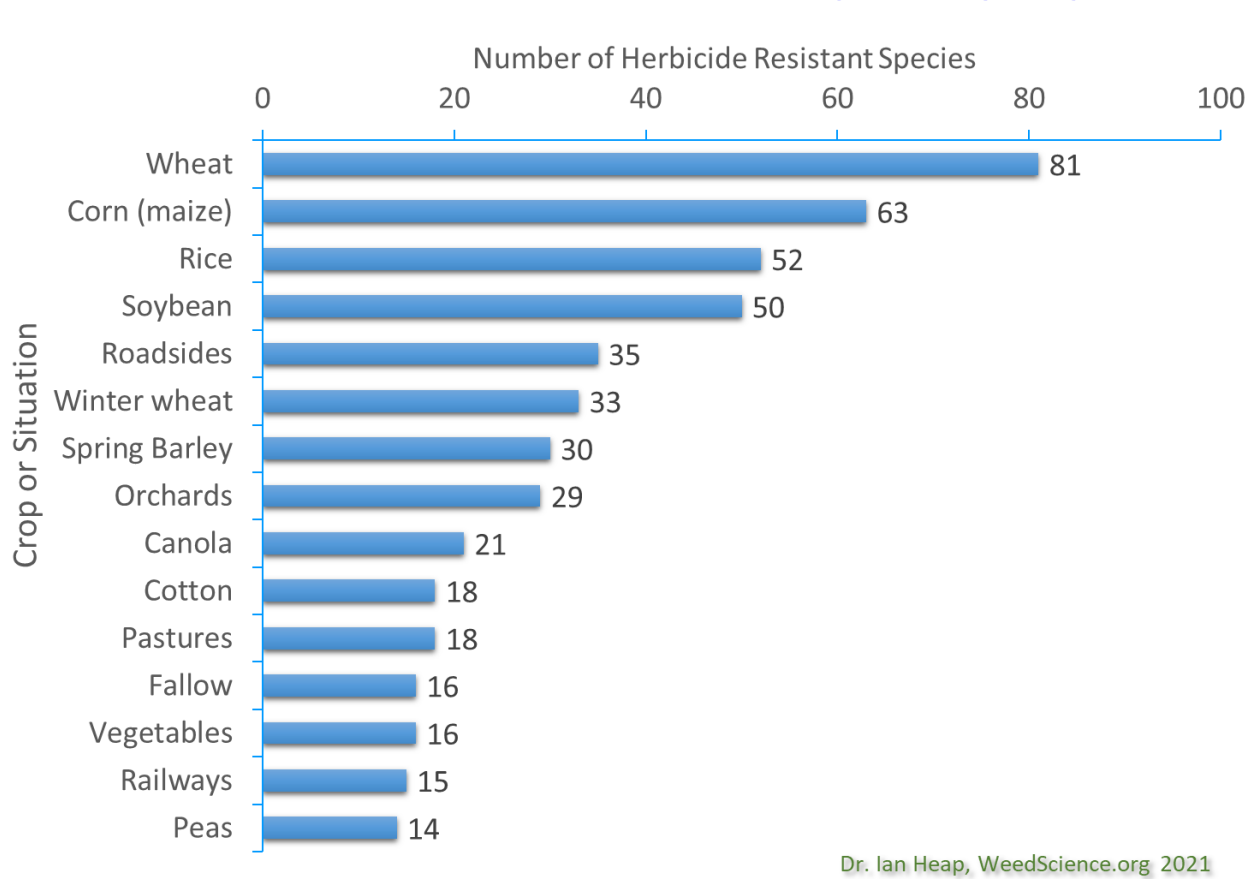
- *Avena* spp.
- *Bromus* spp.
- *Erigeron* spp.
- *Hordeum leporinum*
- *Lolium rigidum*
- *Raphanus raphanistrum*
- *Sisymbrium orientale*
- *Sonchus serriola*

Sources:
Regional HRAC (EU, US, ASIA, JP, BR, ARG, ZA, AUS)

* Indian + Pakistan

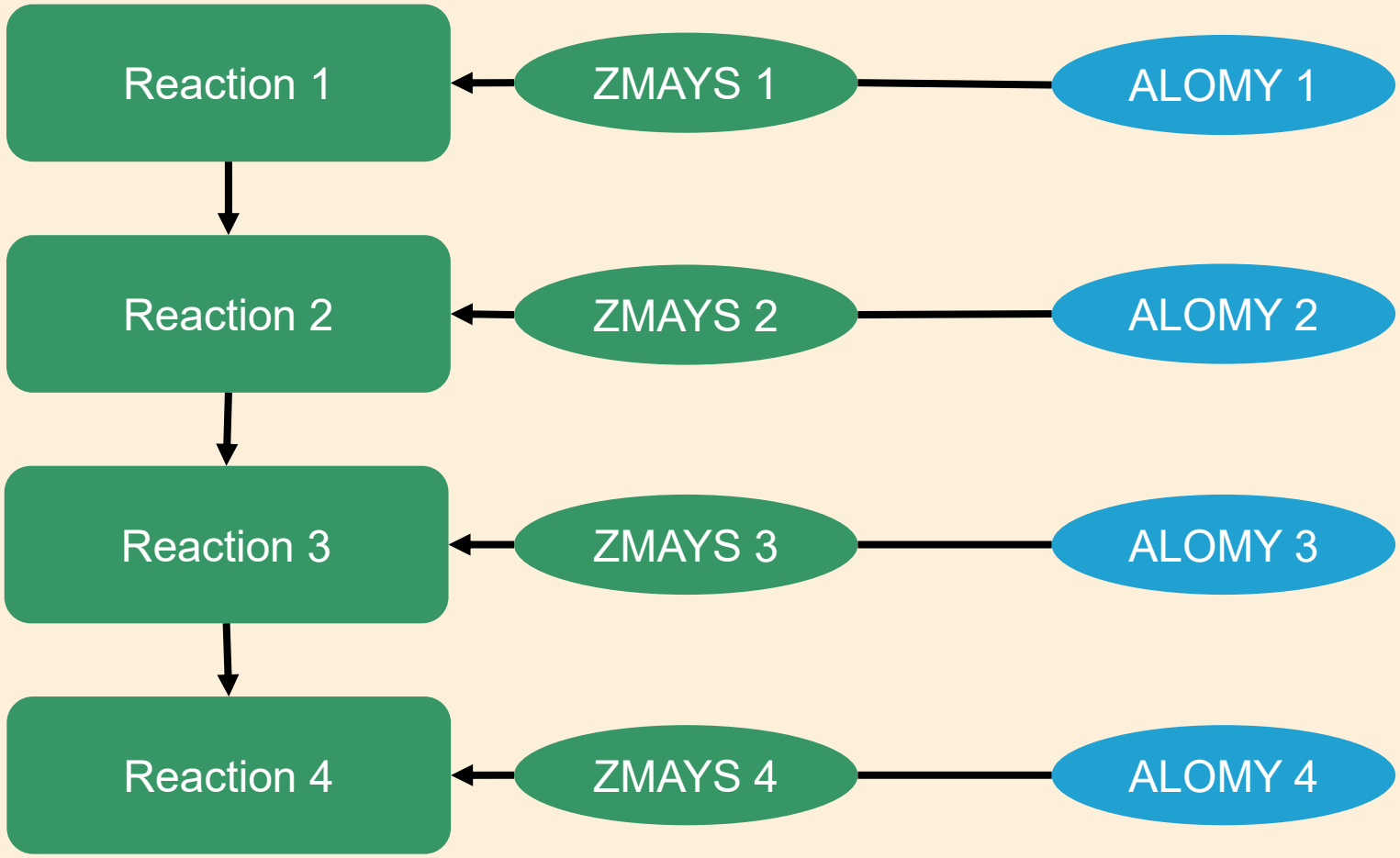
Pests Adapt to Management - Management needs to Adapt too!



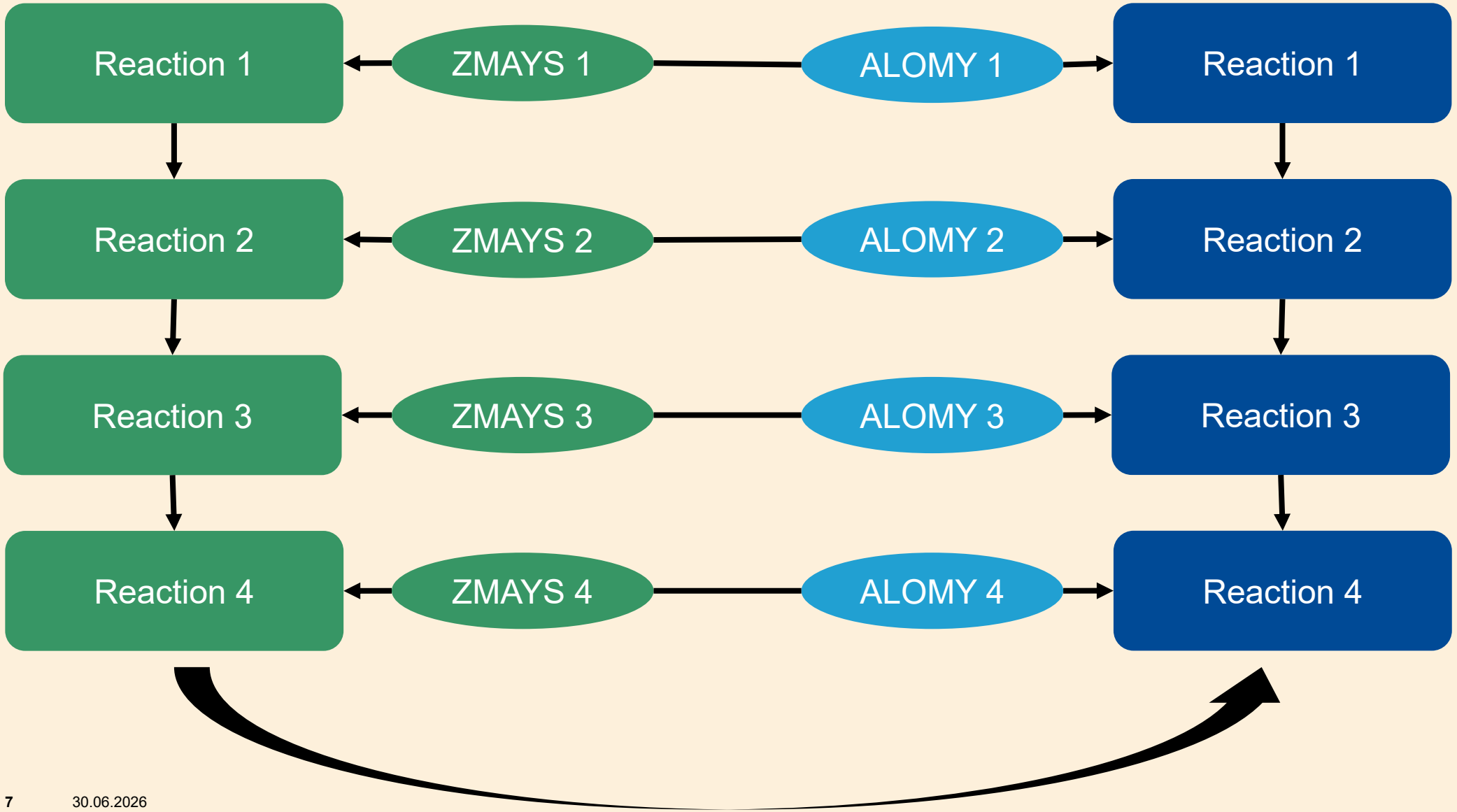


BASF
We create chemistry

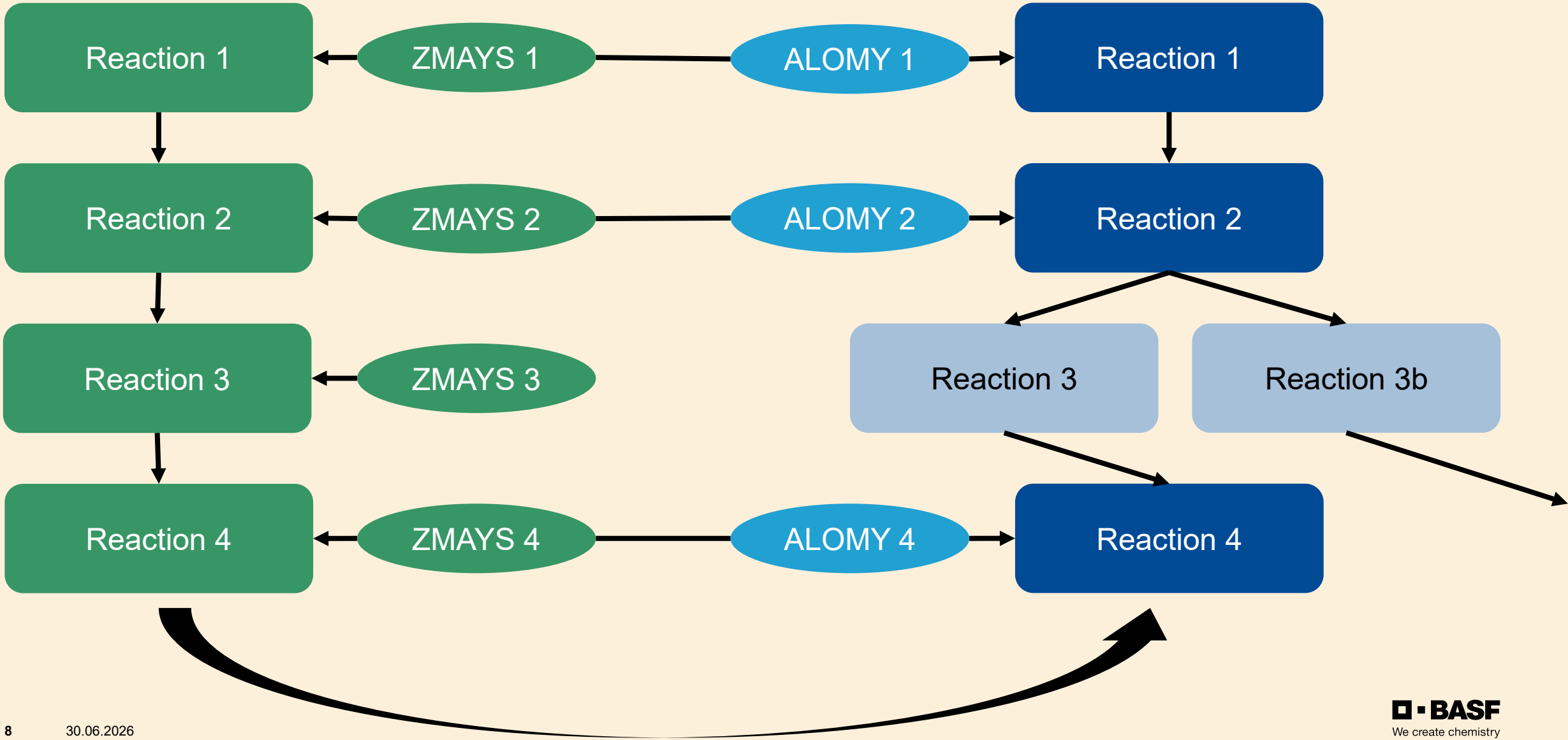
Orthology



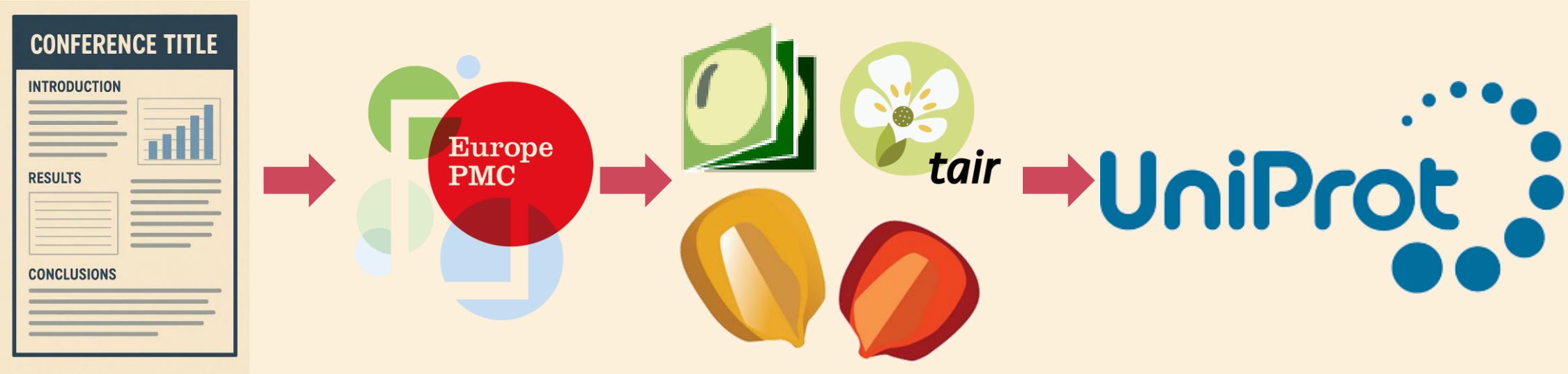
Orthology



Orthology



Knowledge Flows into General Databases Over Time



Quality

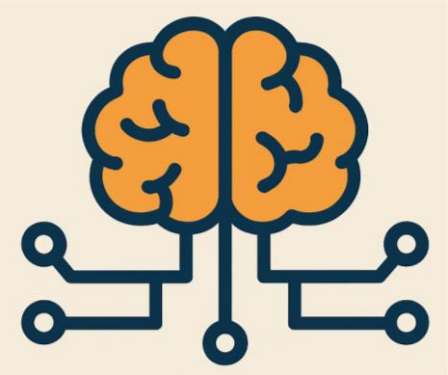
Quantity & Specificity

BASF
We create chemistry

Need: Consolidate Biological Context from knowledge-rich systems to target knowledge-poor systems

- Status quo: standard annotation tools provide links to databases & short descriptions
- Integrate knowledge enhances representation of non-model systems & hypothesis generation from experimental data

Facilitate Data Integration

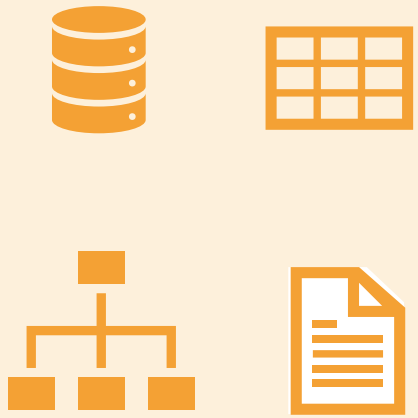


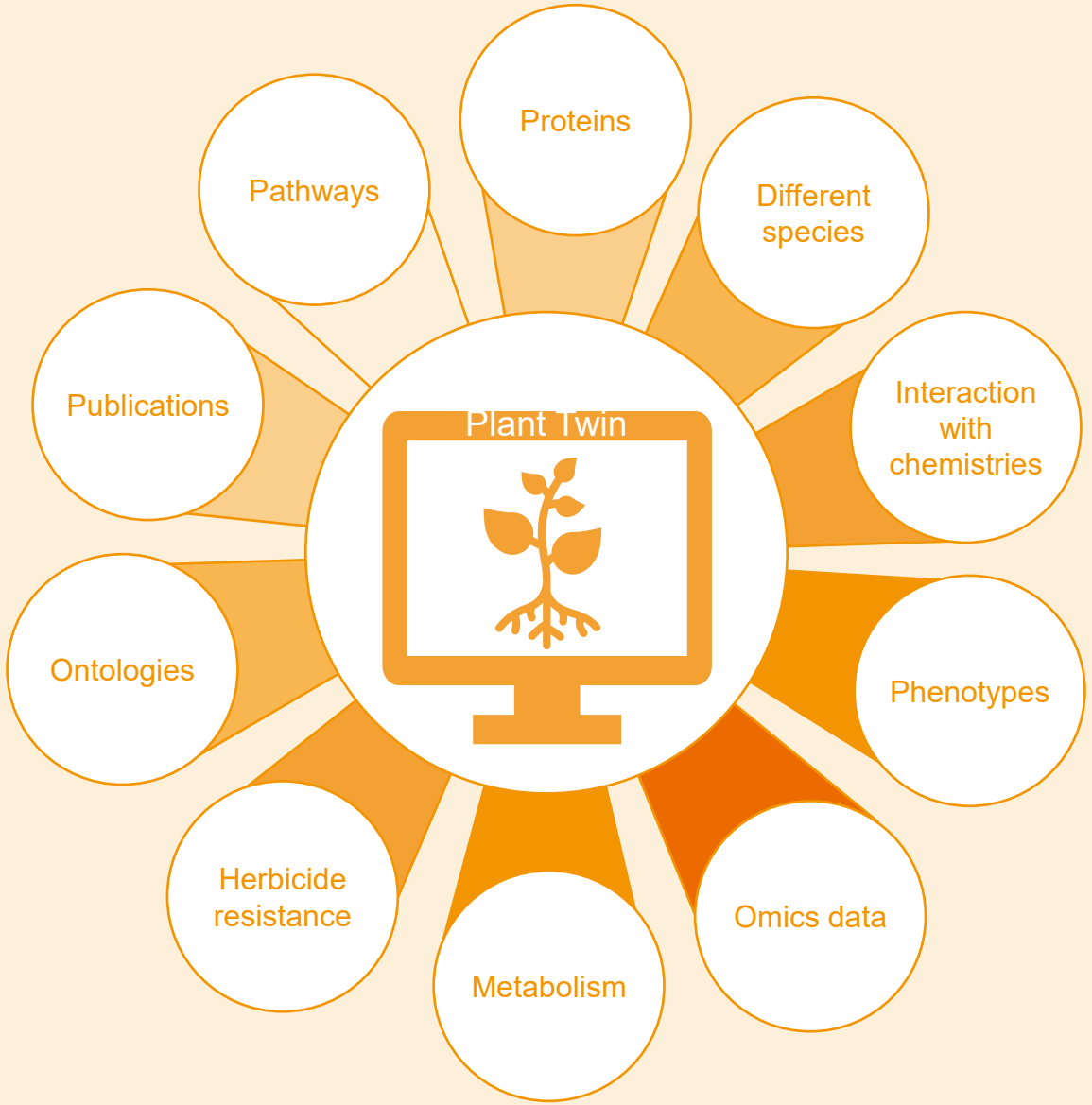
Improve Data AI Readiness

Make data FAIR

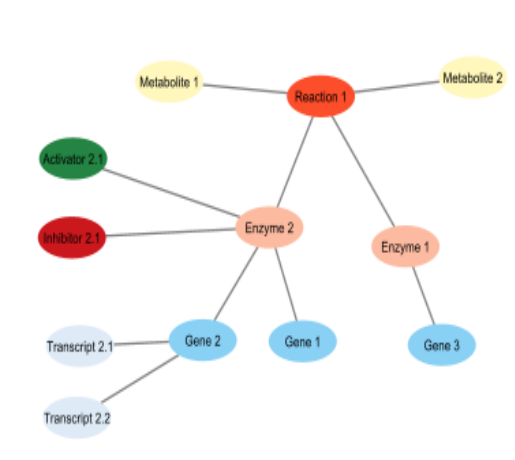
Solution: Create a „Plant Twin“ Graph

Sources of information



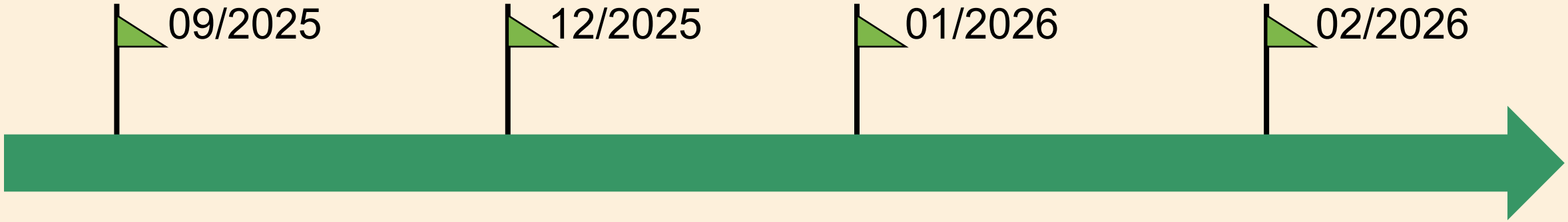


Neo4j Graph db + UI as the technical solution



BASF
We create chemistry

Process



Project Initiation
Dataset identification
Data model ideation

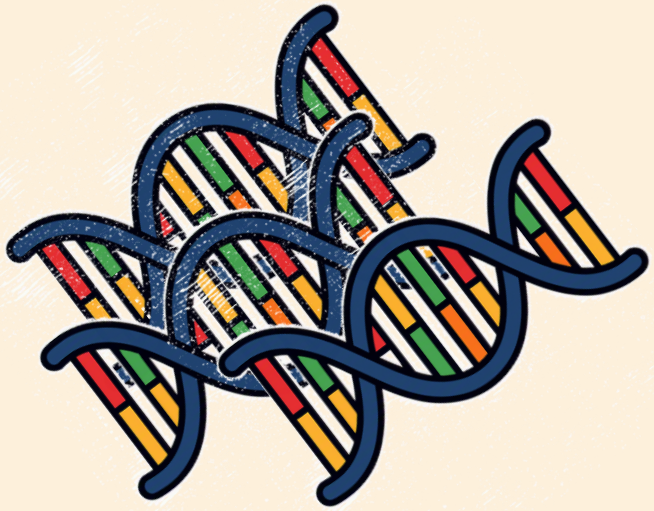
Dataset integration
Initial use-case evaluation

Dataset integration
Data model refinement
IPOV integration into agentic systems

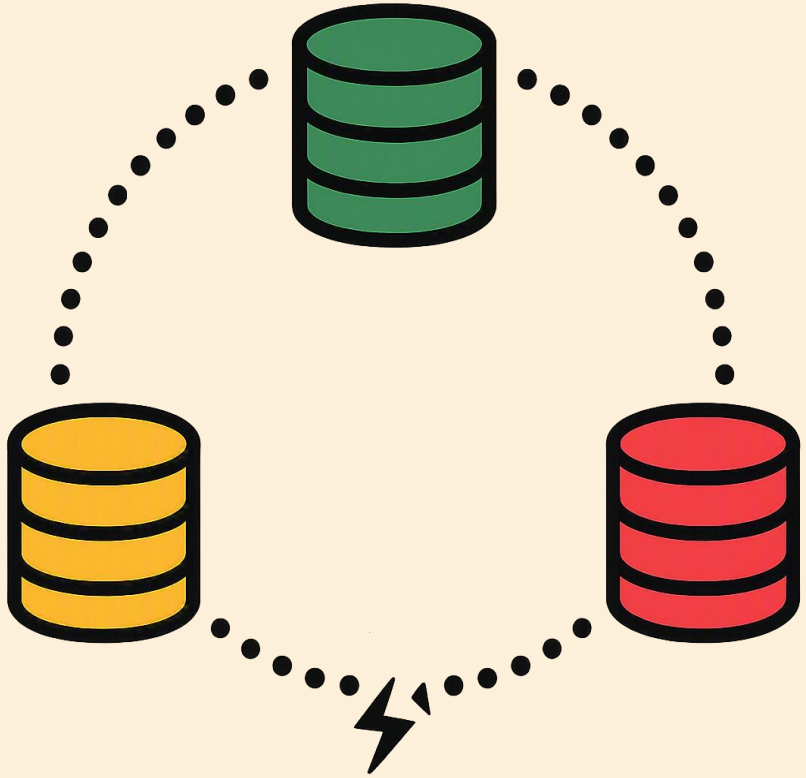
Use-case evaluation
POV defense
Ideation & next steps

- Collaboration with Neo4j Professional Services
- Went from nothing to successful POV over a 6-month timeframe, ~3 “person months”
- Main challenges:
  - Which sources (compare GEMs, PPI)
  - Identifier mappings (external & internal identifiers)

Data Integration is non-Trivial

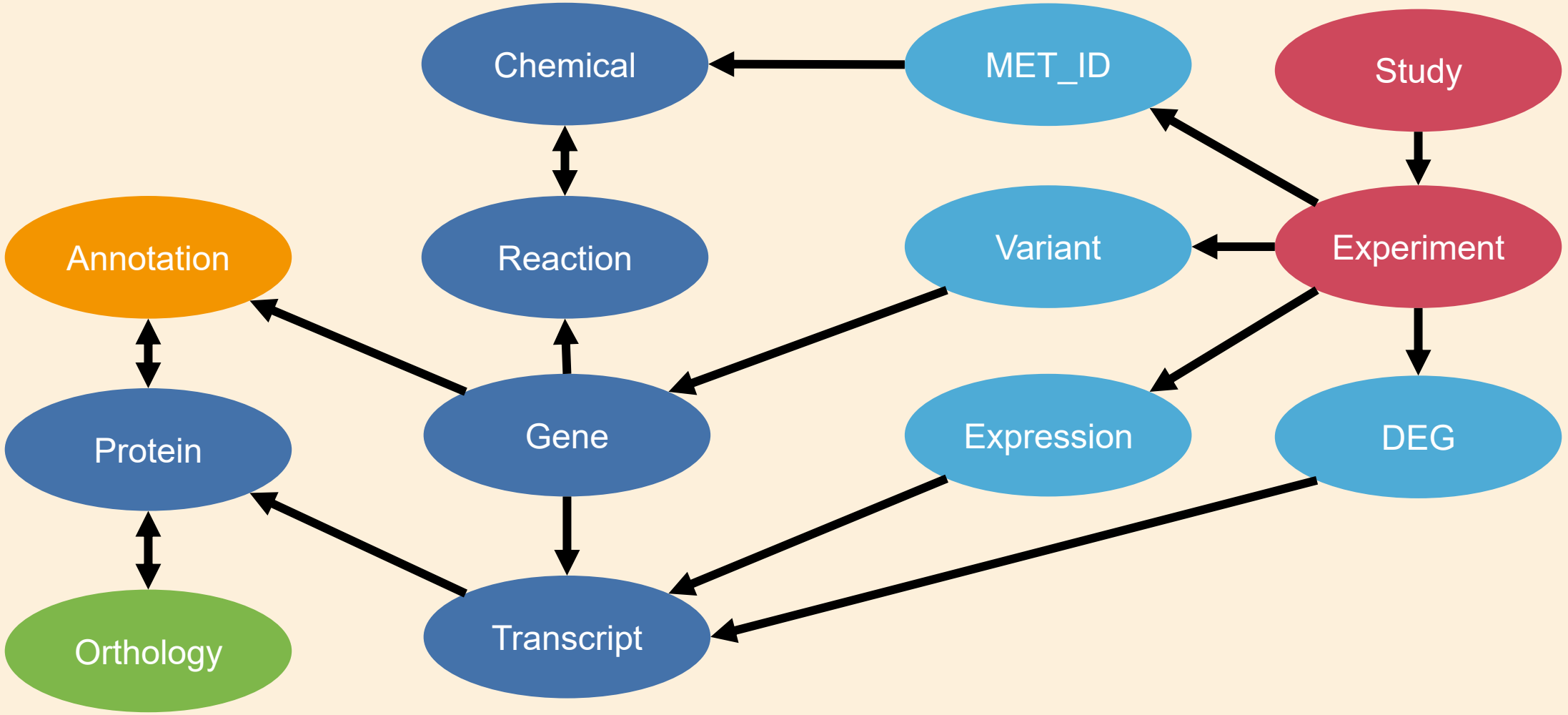


Genome
versions
Gene IDs



Identifier
compatibility

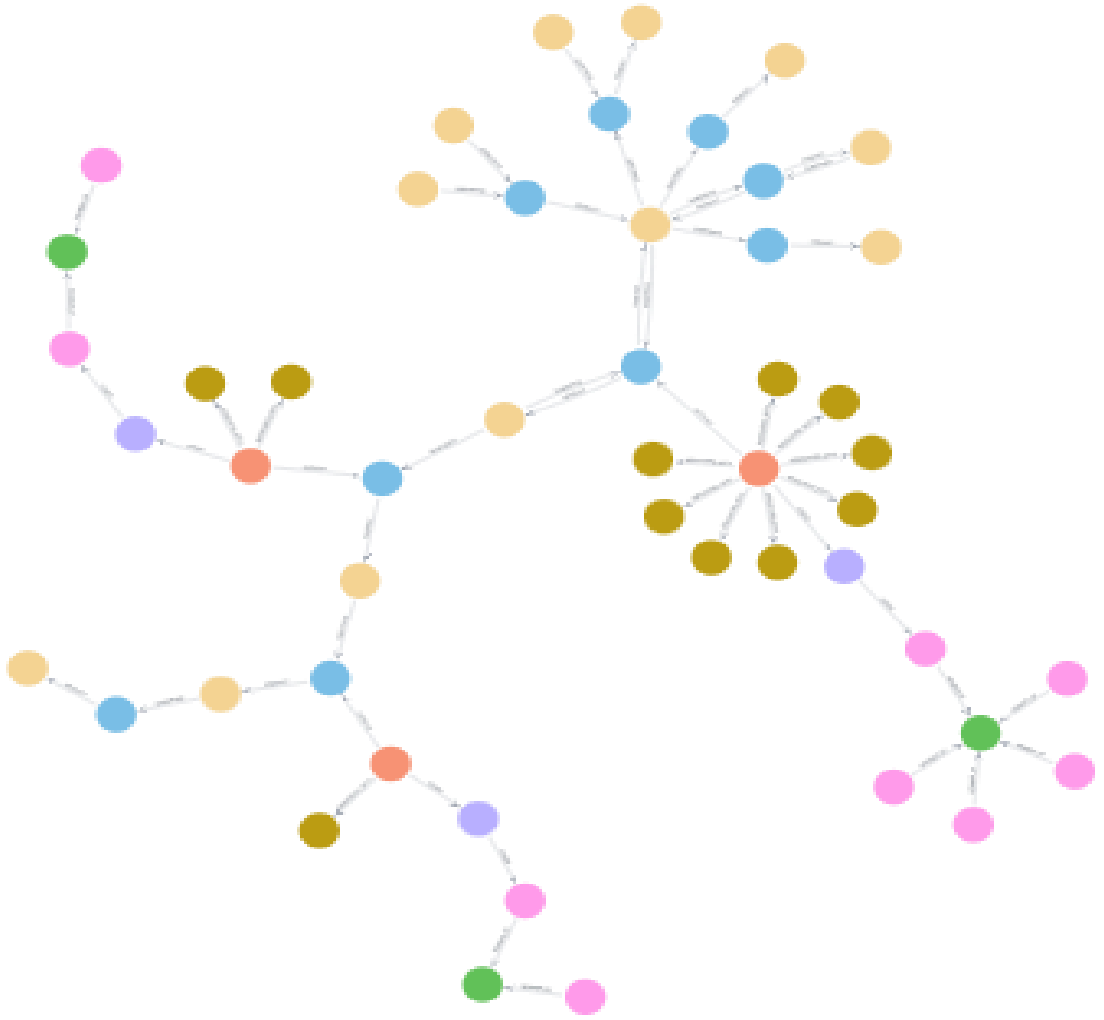
Data model: core graph



Example: Dwarfing genes in Lemna spp

- 5 datasets

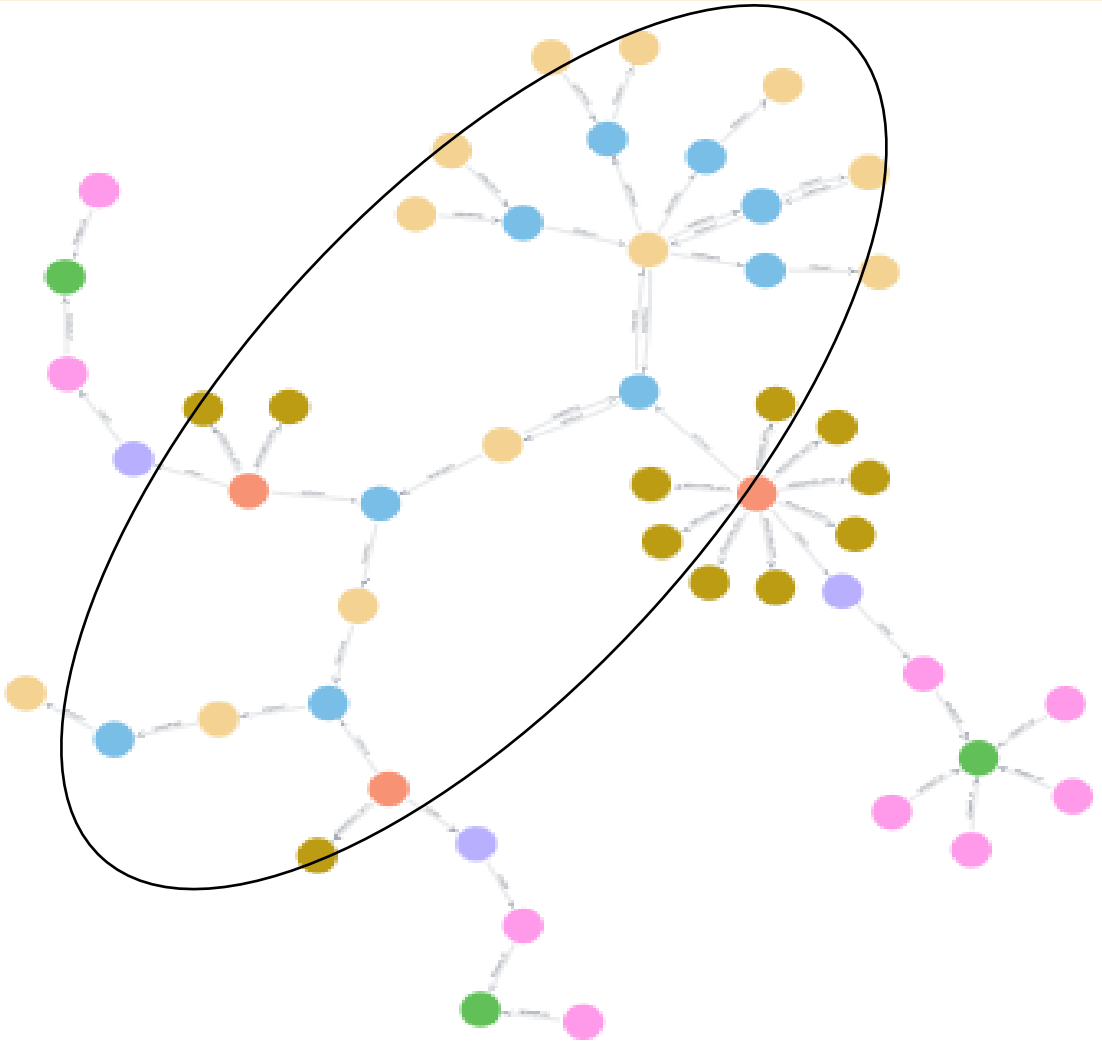
Phenotype
Gene
Protein
OrthoGroup
Reaction
Chemical



Example: Dwarfing genes in Lemna spp

- 5 datasets
- Reaction network

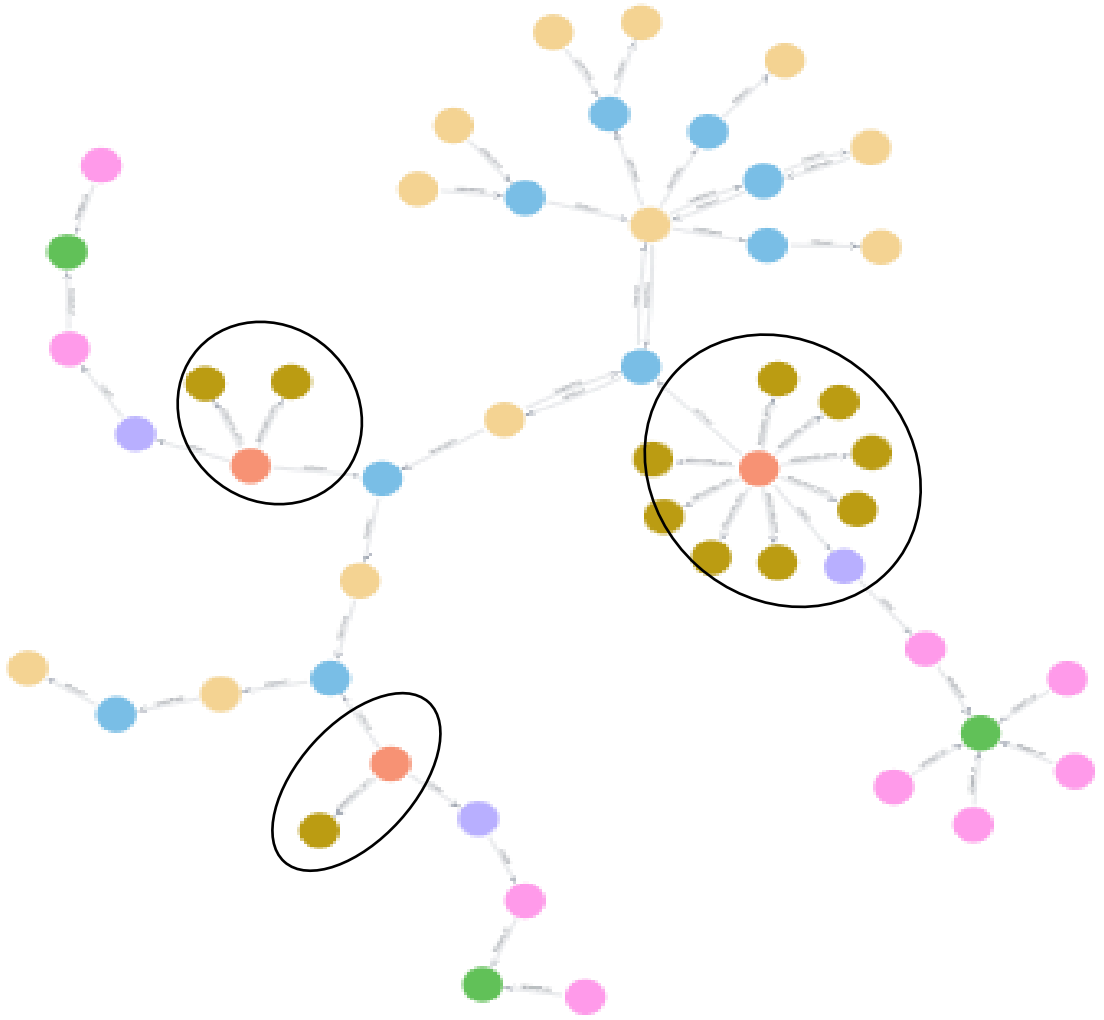
Phenotype
Gene
Protein
OrthoGroup
Reaction
Chemical



Example: Dwarfing genes in Lemna spp

- 5 datasets
- Reaction network
- Literature-derived phenotypes

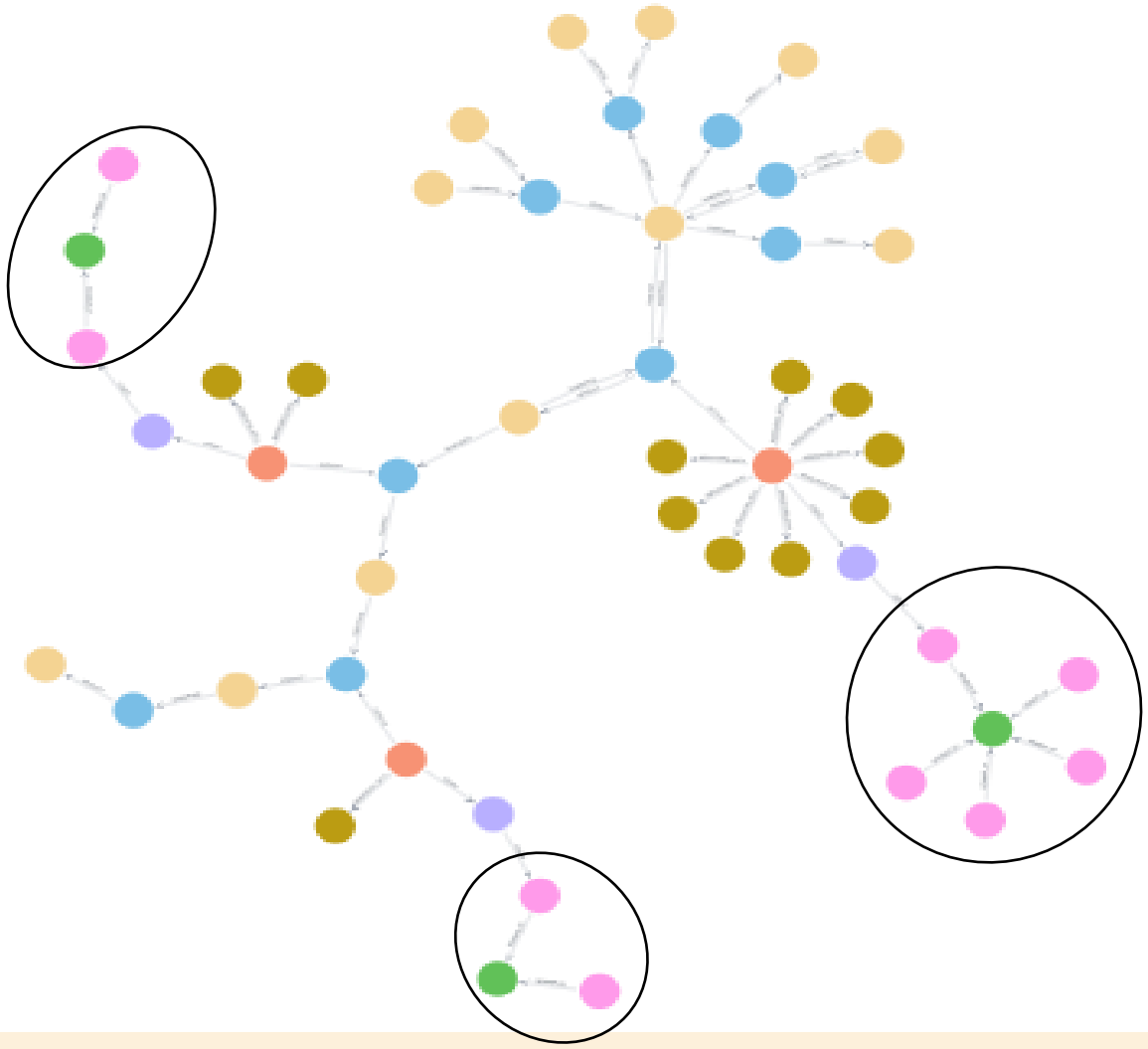
Phenotype
Gene
Protein
OrthoGroup
Reaction
Chemical



Example: Dwarfing genes in Lemna spp

- 5 datasets
- Reaction network
- Literature-derived phenotypes
- Orthology groups

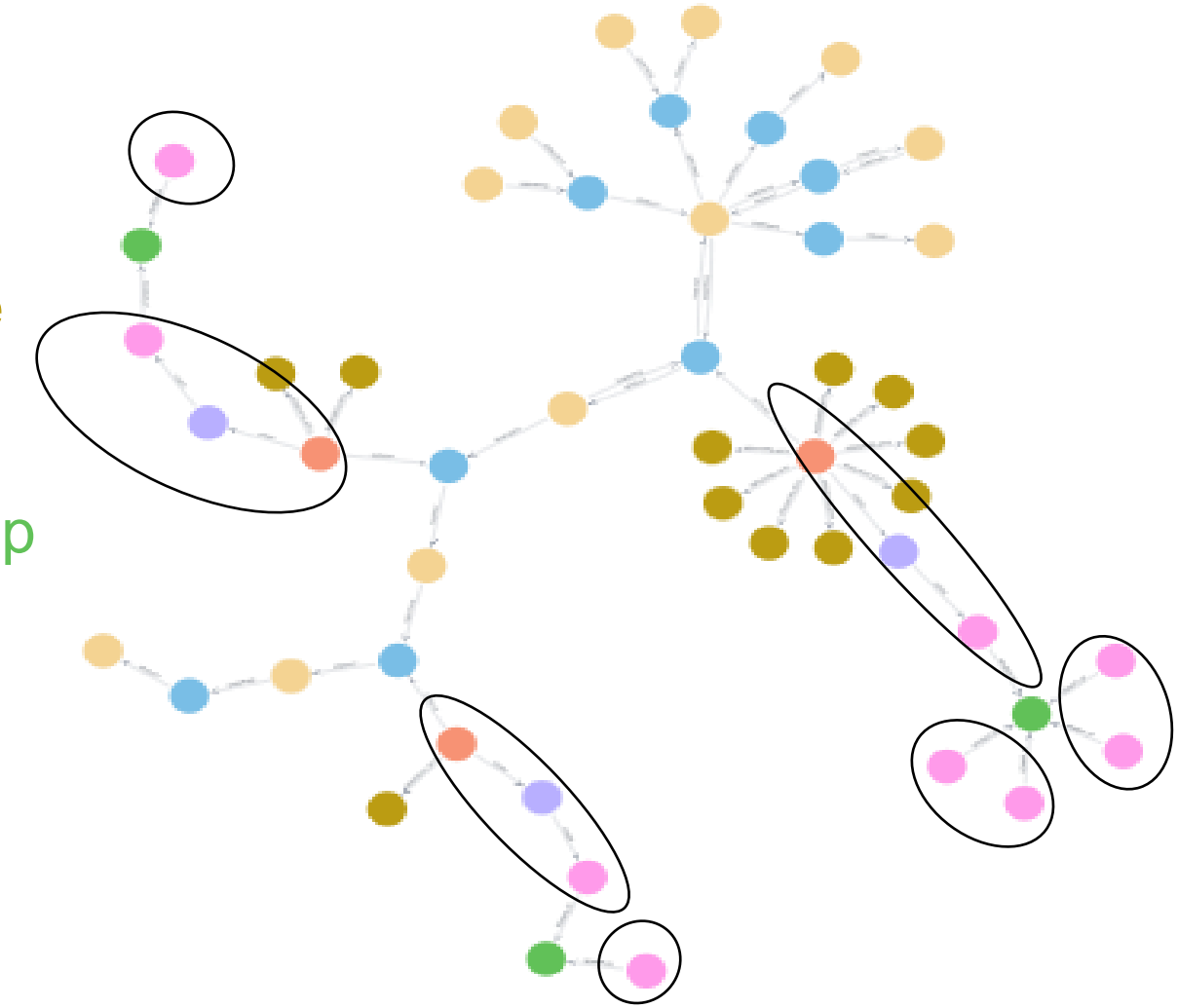
Phenotype
Gene
Protein
OrthoGroup
Reaction
Chemical



Example: Dwarfing genes in Lemna spp

- 5 datasets
- Reaction network
- Literature-derived phenotypes
- Orthology groups
- ARATH & Lemna genomes

Phenotype
Gene
Protein
OrthoGroup
Reaction
Chemical



Annotation task: what does 25940:g021888 do?

- Graph-driven AI summary: use orthogroup to supplement annotations

25940:g021888 is most likely a class II diterpene synthase, best annotated as a probable ent-copalyl diphosphate synthase / copalyl diphosphate synthase-like protein.
In functional terms, this places it in the **early steps of diterpenoid biosynthesis**, likely generating a cyclized diphosphate intermediate used for **gibberellin biosynthesis and/or specialized diterpene metabolism**.

- Interproscan:

Terpenoid cyclases/protein prenyltransferase alpha-alpha toroid; ENT-COPALYL DIPHOSPHATE SYNTHASE, CHLOROPLASTIC; pfam: Terpene synthase family, metal binding domain

- Eggnog

Desc: Belongs to the terpene synthase family; EC: 4.2.3.19; pfams: Terpene_synth, Terpene_synth_C

Annotation task: what does 25940:g021888 do?

- AI-driven graph follow-up (also use PPI & reaction network information)?

is probably not just any terpene synthase-like protein; it is most likely a plastidial class II diterpene cyclase that converts GGPP into ent-copalyl diphosphate or a very closely related intermediate, placing it at the entry point to ent-kaurene/gibberellin biosynthesis and possibly related specialized diterpene branches

Acknowledgements

- Neo4j Consultancy Team: Martin Preusse, Stuart Green, Anastasia Puchkova
- BASF Core Team: Brent Murphy, Averina Nicolae, Vidya Oruganti
- Agentic systems: Eliot Boulanger, Koen Bruynseels
- Software & Infrastructure Support: Carsten Seidel, Bas Van Criel, Kristof Egghe, Thibault Gobert, Daniel Rehm, Juergen Geschwill

We create chemistry