WFCC ICCC-12 Conference 2010

WORLD FEDERATION FOR CULTURE COLLECTIONS

Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

EVOLUTION OF THE YEAST SPECIES CONCEPT IN THE AGE OF SEQUENCING

Marc-André Lachance

Department of Biology
University of Western Ontario
London, Ontario, Canada N6A 5B7
lachance@uwo.ca

Western Biology

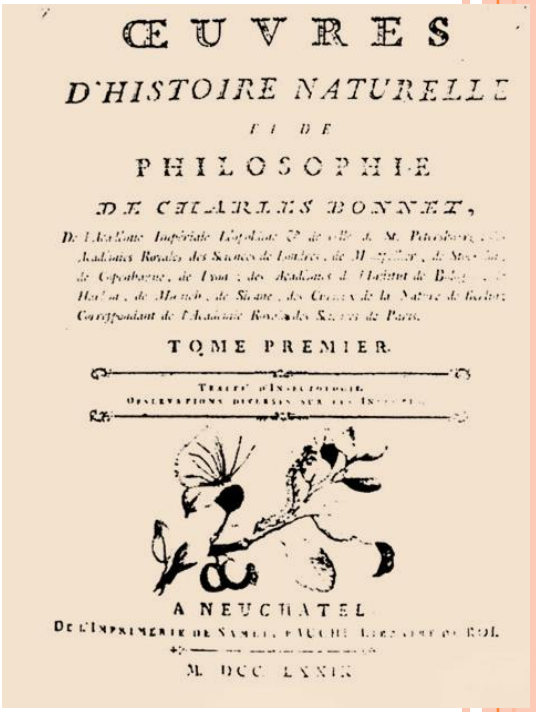
EXPLORATION OF YEAST BIODIVERSITY REQUIRES A SOUND SPECIES CONCEPT

Classis et ordo est sapientiae, species naturae opus

Classe e ordem são o trabalho da ciência, a espécie o da natureza

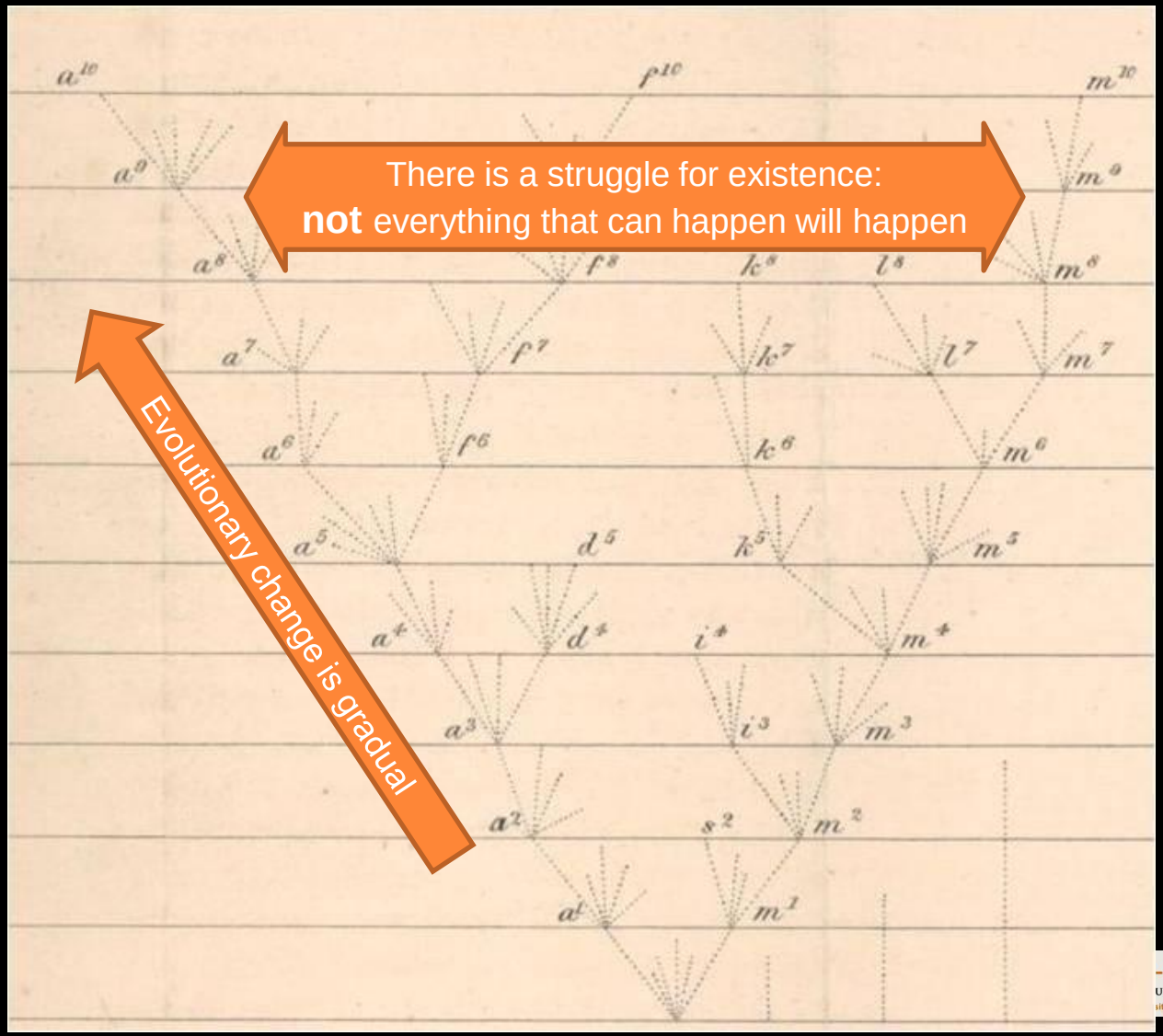
METAPHYSICS OF SPECIES

- Principle of plenitude (Aristotle)
  - Everything that can happen will happen
  - The tree of life will fill up with every possible clade
- Principle of continuity (Aristotle, Linnaeus, Leibnitz, Newton, La marck, Darwin)
  - Natura non facit saltum
  - Evolutionary change is gradual



Western Biology
WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

ENTERS DARWIN



SPECIES CONCEPTS

- Biological Species Concept
  - "Groups of interbreeding natural populations that are reproductively isolated from other such groups" (Mayr)
    - Unit of evolution
    - An intrinsic property of the members (objective)
    - Testable/Falsifiable (*sensu* Popper)
    - Limited to heterothallic, sexual species

SPECIES CONCEPTS

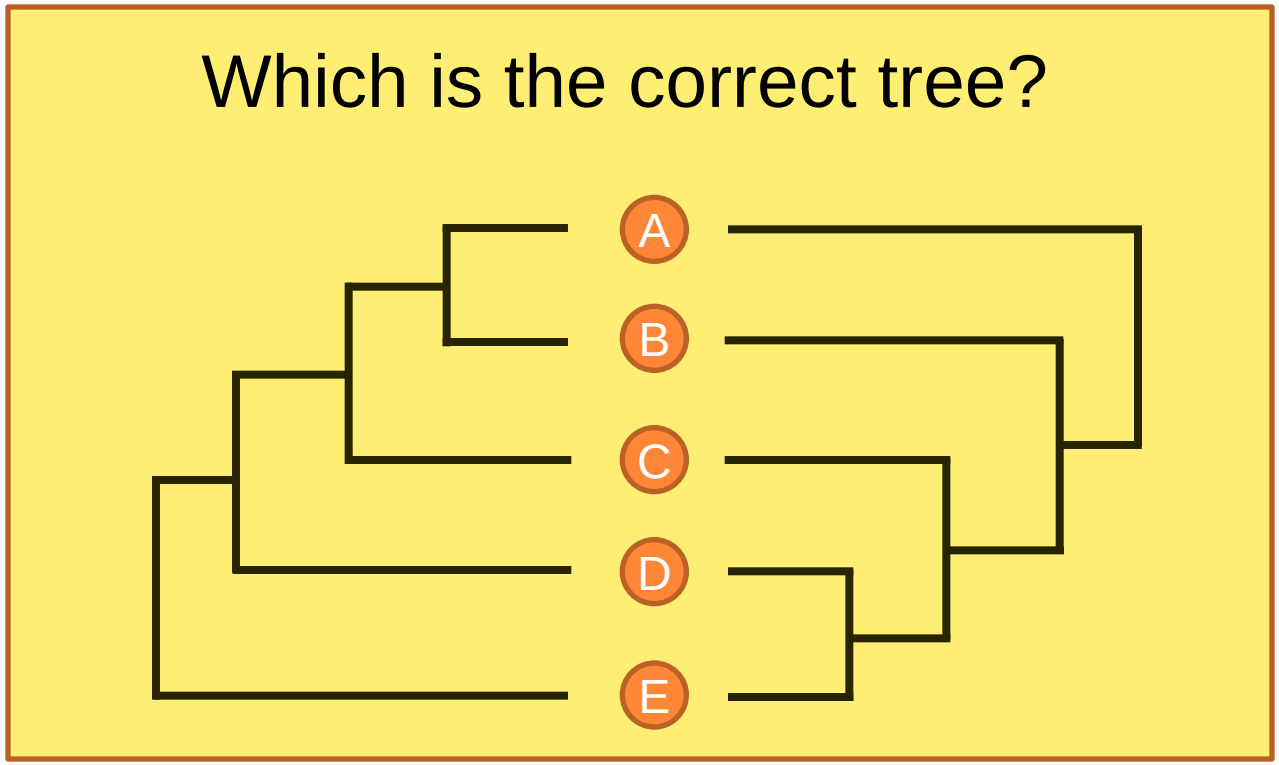
- Genetic Species Concept
  - "Groups of individuals that share a high amount of genetic similarity and are distinct from other such groups"
    - Microbiological version uses DNA/DNA reassociation
      - 80% for yeasts
    - Calibrated on the Biological Species Concept (proxy)
  - Broadly applicable
  - Not always testable (arbitrary)
  - Onerous… sequencing is a lot easier!

SPECIES CONCEPTS

- Phylogenetic Species Concept
  - "The least inclusive monophyletic assemblage that can be recognized from a unique set of characteristics"
    - Can be based on DNA sequence analyses
    - Reflects the history of the members (objective)
    - Rarely testable (*sensu* Popper)
    - Inclusivity difficult to define

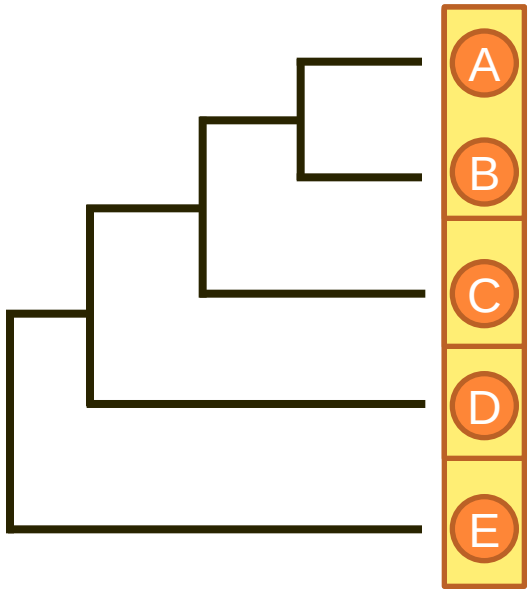
EPISTEMOLOGY: PHYLOGENETIC CONCEPT

1. A tree is always an inference



EPISTEMOLOGY: PHYLOGENETIC CONCEPT

2. Assuming that this is the correct tree, which is the least inclusive assemblage?



The amount of differences considered necessary to give to any two forms the rank of species cannot be defined.

Darwin 1859

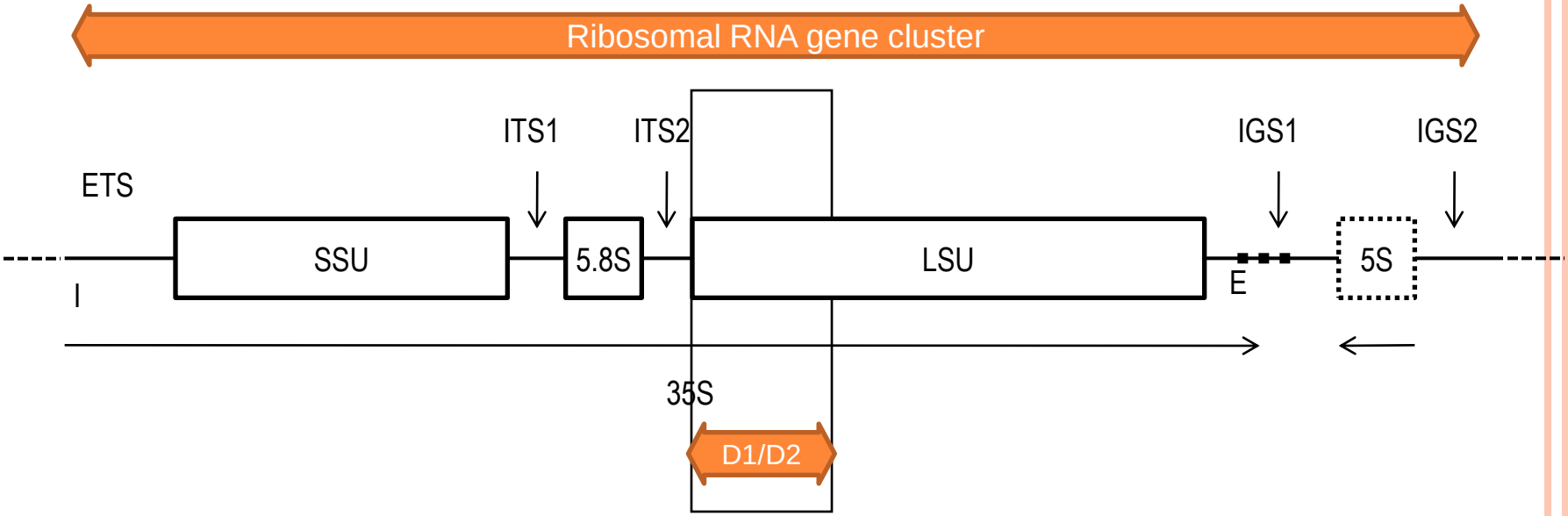
WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

SPECIES CONCEPTS

- *Sequence Divergence* Species Concept
  - "Groups of individuals that differ by 3 or fewer substitutions in the D1/D2 LSU rRNA gene and differ from other such groups by 1% or more divergence
    - Can be based on DNA sequence analyses
    - Broadly applicable
      - Some notorious exceptions
    - Not always testable (arbitrary)

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

YEAST BARCODING SEQUENCE



WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

SPECIES CONCEPTS

- *Sequence Divergence* Species Concept
  - "Groups of individuals that differ by 3 or fewer substitutions in the D1/D2 LSU rRNA gene and differ from other such groups by 1% or more divergence
    - Can be based on DNA sequence analyses
    - Broadly applicable
      - Some notorious exceptions
    - Not always testable (arbitrary)

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

SEQUENCE DIVERGENCE CONCEPT - EXCEPTIONS

- *Metschnikowia agaves*
  - Holotype and allotype differ by 5 substitutions
    - Kurtzman & Robnett 1998
- *Clavispora lusitaniae*
  - Holotype *et al* differ by > 30 substitutions from allotype *et al*
  - Some strains contain both D2 variants
    - Lachance et al. 2003

SPECIES CONCEPTS

- *Sequence Divergence* Species Concept
  - "Groups of individuals that differ by 3 or fewer substitutions in the D1/D2 LSU rRNA gene and differ from other such groups by 1% or more divergence
    - Can be based on DNA sequence analyses
    - Broadly applicable
      - Some notorious exceptions
    - Not always testable (arbitrary)

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

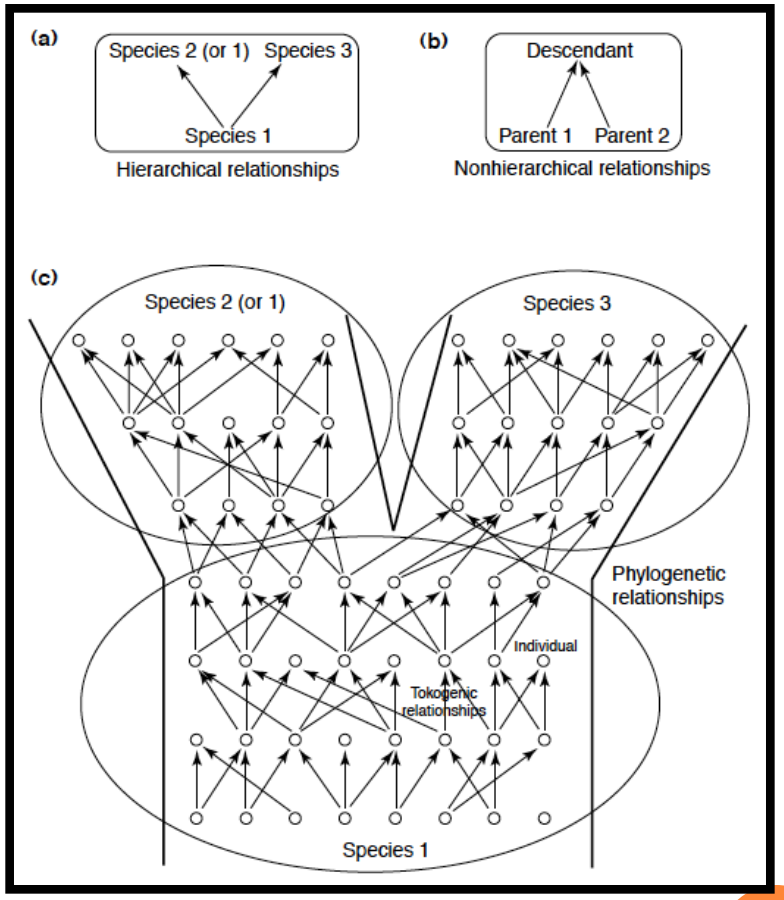
SEQUENCE DIVERGENCE CONCEPT - EPISTEMOLOGY

- Induction
  - "Taking into account the variation seen in the preceding comparisons, it is *predicted* that strains showing greater than 1% substitutions in the ca. 600-nucleotide D1/D2 domain *are likely to be different species* and that strains with 0–3 nucleotide differences are either conspecific or sister species." (Kurtzman and Robnett 1998)
- Generalization
  - "Kurtzman and Robnett (1998) *demonstrated* for ascomycetous yeasts that strains differing by more than 1% substitutions in the D1/D2 domain *represent separate species*." (Kurtzman and Droby 2001)

TOKOGENY *VS* PHYLOGENY

"The only distinction between species and well-marked **varieties** is, that the latter **are** known, or believed, to be **connected at the present day** by intermediate gradations, whereas **species were formerly thus connected**."

Darwin 1859



Posada & Crandall 2001

ex Hennig 1966

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

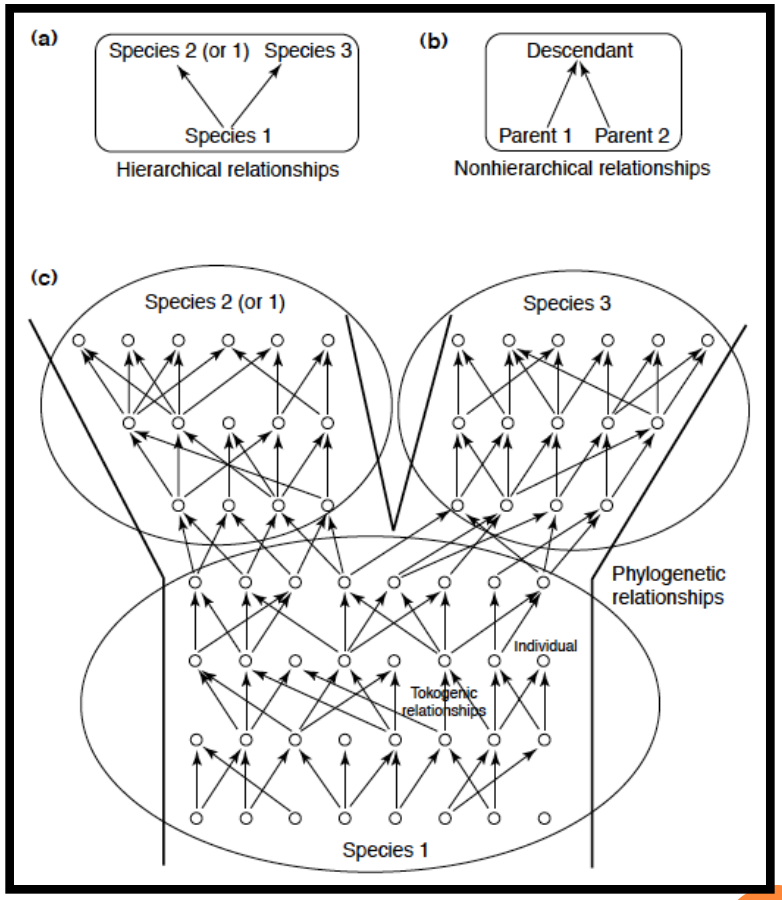
TOKOGENY *VS* PHYLOGENY

“The only distinction
be
ma ne
latt
be
at the present day by
int
gra
sp us
co

Varieties are connected *at the present day*

species were *formerly* thus connected

Darwin 1859



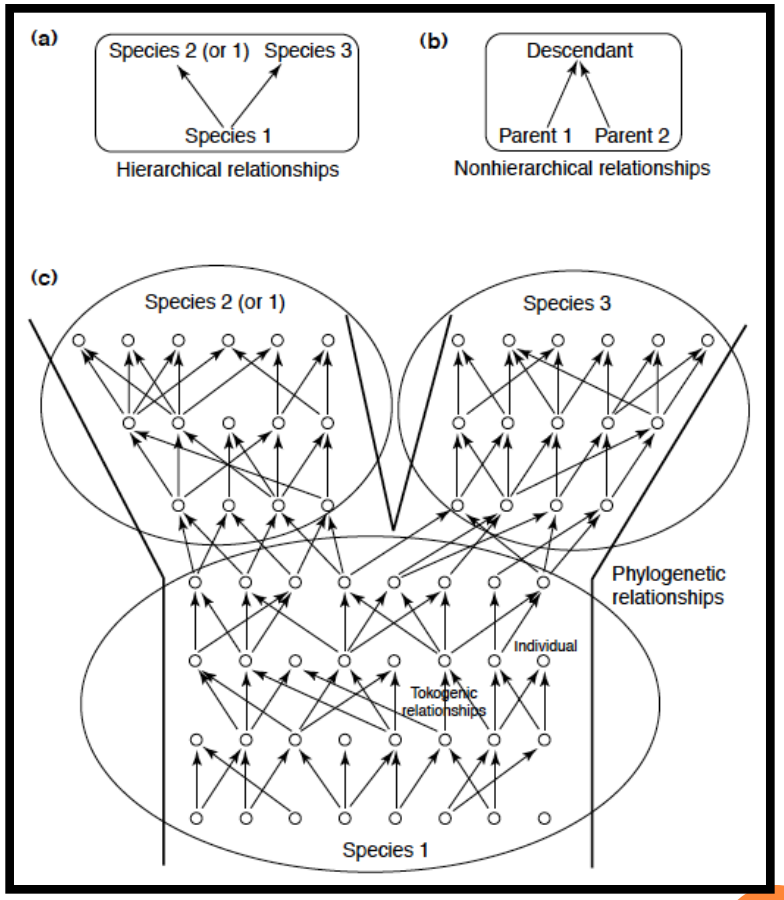
Posada & Crandall 2001

ex Hennig 1966

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

TOKOGENY *VS* PHYLOGENY

- Sequence relationships *between* species are phylogenetic, hierarchic al,
  i.e., **tree-like**
- Sequence relationships *within* species are tokogenetic, nonhierarc hical,
  i.e., **network-like**

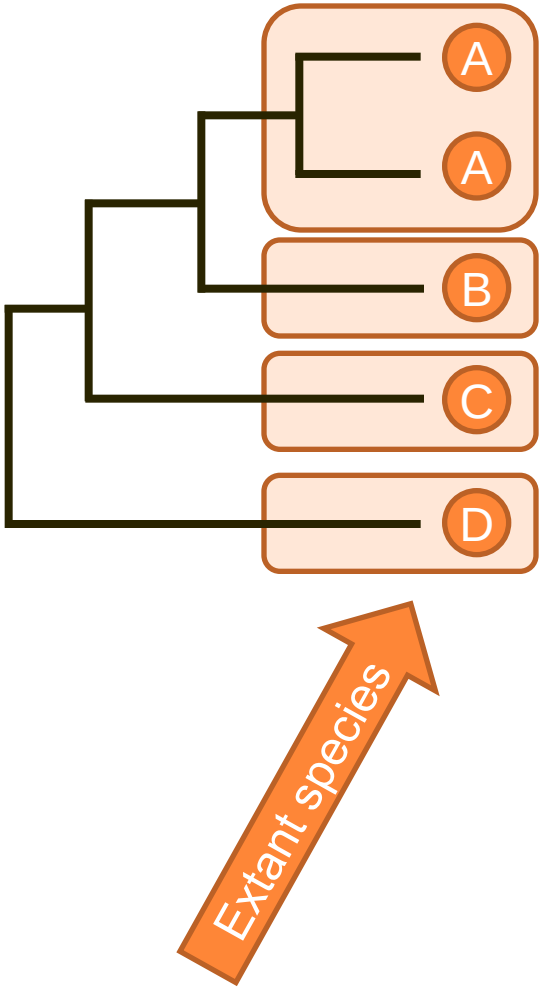


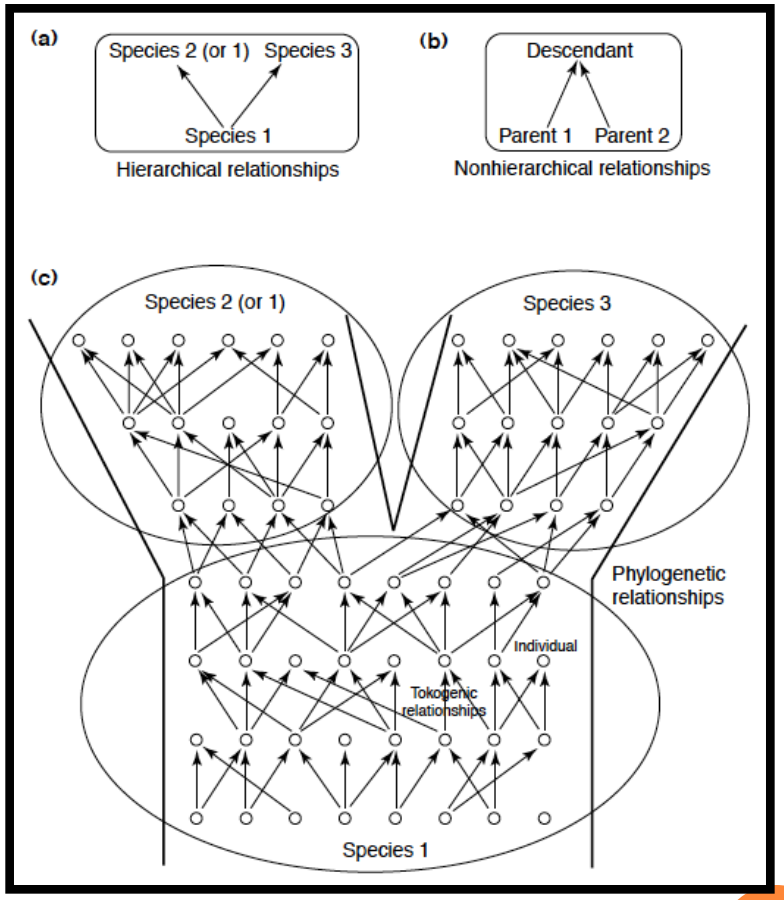
Posada & Crandall 2001

ex Hennig 1966

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

TOKOGENY *VS* PHYLOGENY



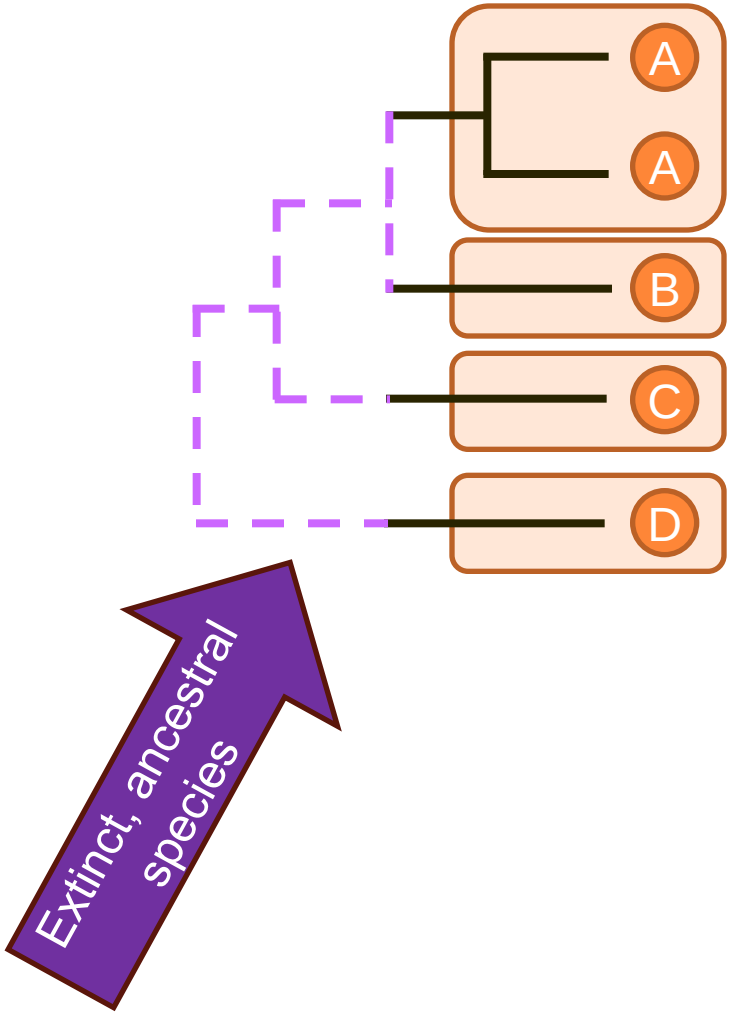


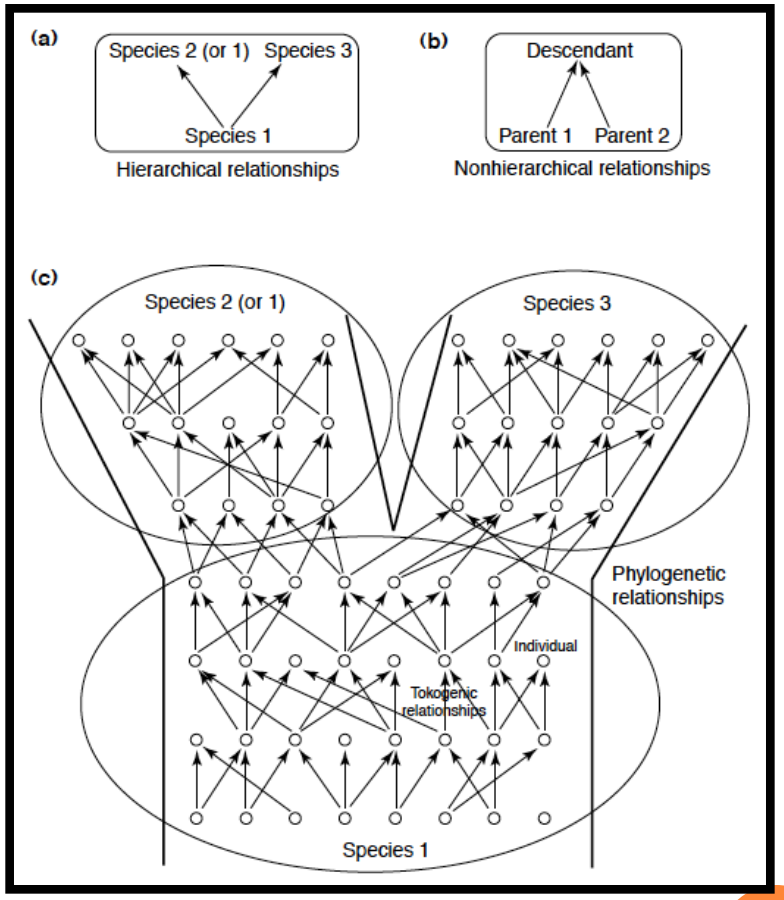
Posada & Crandall 2001

ex Hennig 1966

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

TOKOGENY *VS* PHYLOGENY





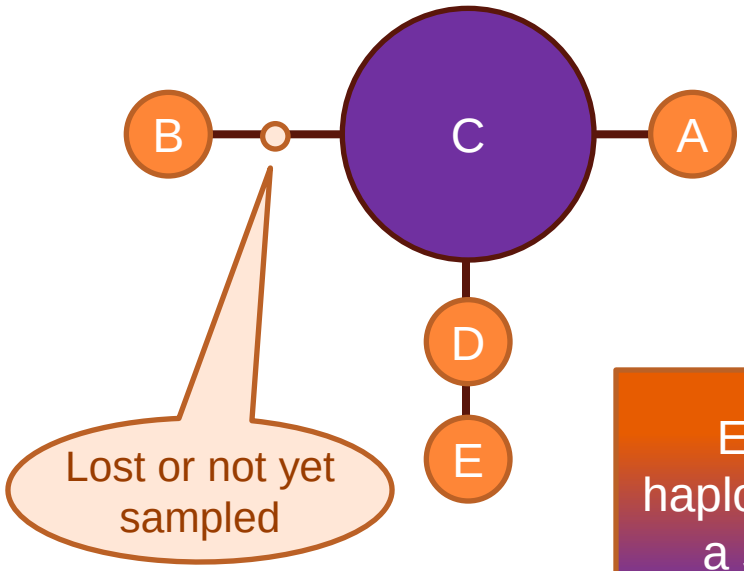
Posada & Crandall 2001

ex Hennig 1966

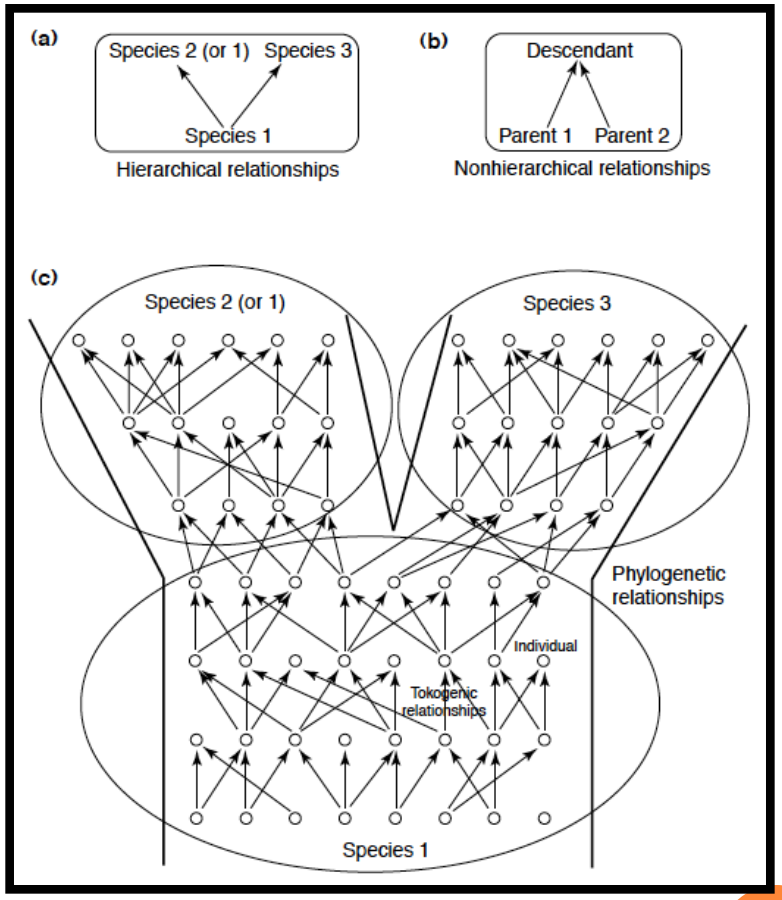
Ancestral sequences no longer exist!

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

TOKOGENY *VS* PHYLOGENY



Extant haplotypes of a single species



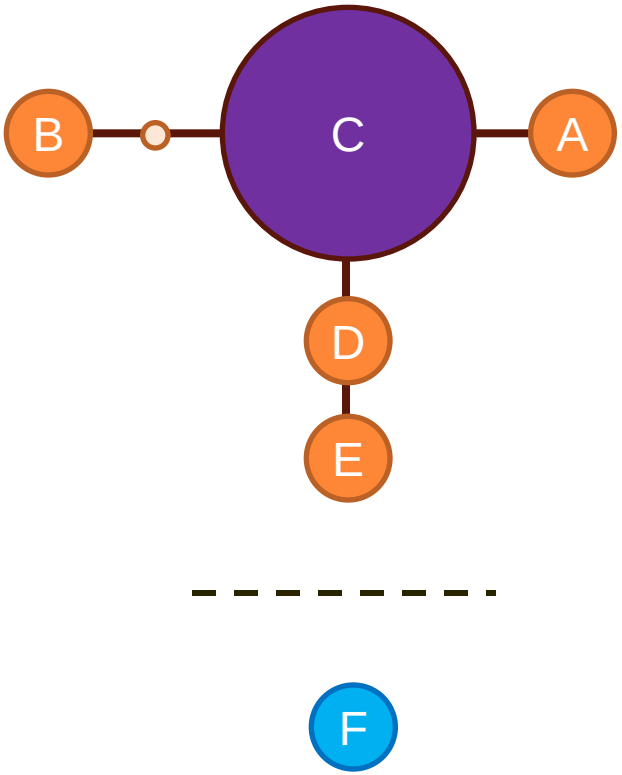
The ancestral sequence is the most abundant!

Posada & Crandall 2001

ex Hennig 1966

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

TOKOGENY *VS* PHYLOGENY

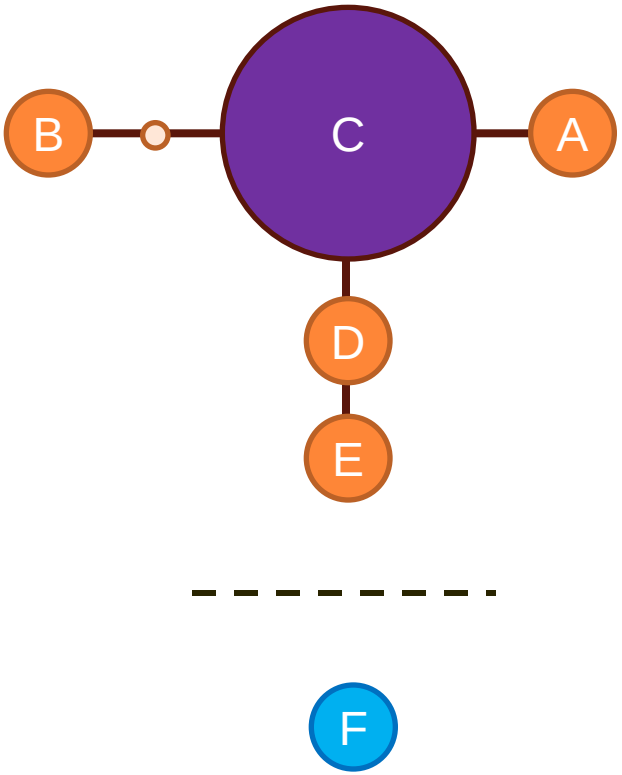


The program **TCS** is an implementation of **T**empleton, **C**randall, and **S**ing's (1992) haplotype network analysis with a statistical parsimony test for membership

Sequences are accepted in a network if the probability that each step represents a single substitution is greater than a specified *P* (0.95)

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

TOKOGENY *VS* PHYLOGENY



biology letters

Biol. Lett. (2007) **3**, 509–512
doi:10.1098/rsbl.2007.0307
Published online 24 July 2007

Evolutionary biology

Things fall apart: biological species form unconnected parsimony networks

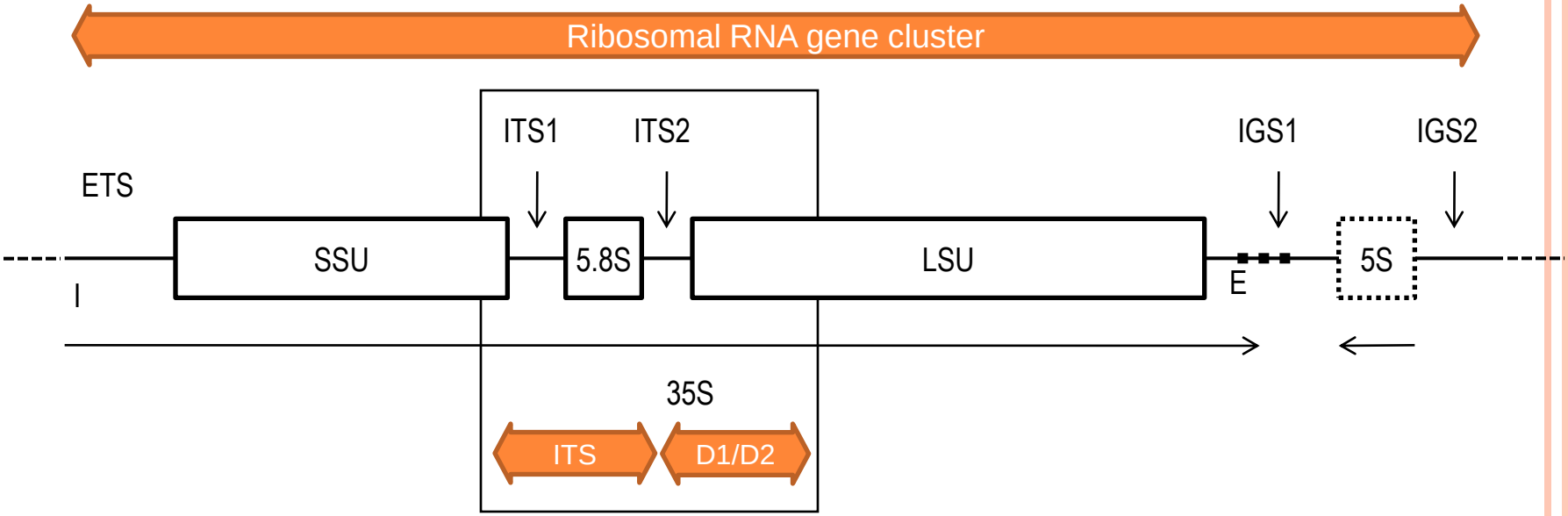
Michael W. Hart* and Jennifer Sunday

Department of Biological Sciences, Simon Fraser University

78% of species studied identified correctly by TCS (*N* = 663 across Eukaryota)

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

YEAST BARCODING SEQUENCE

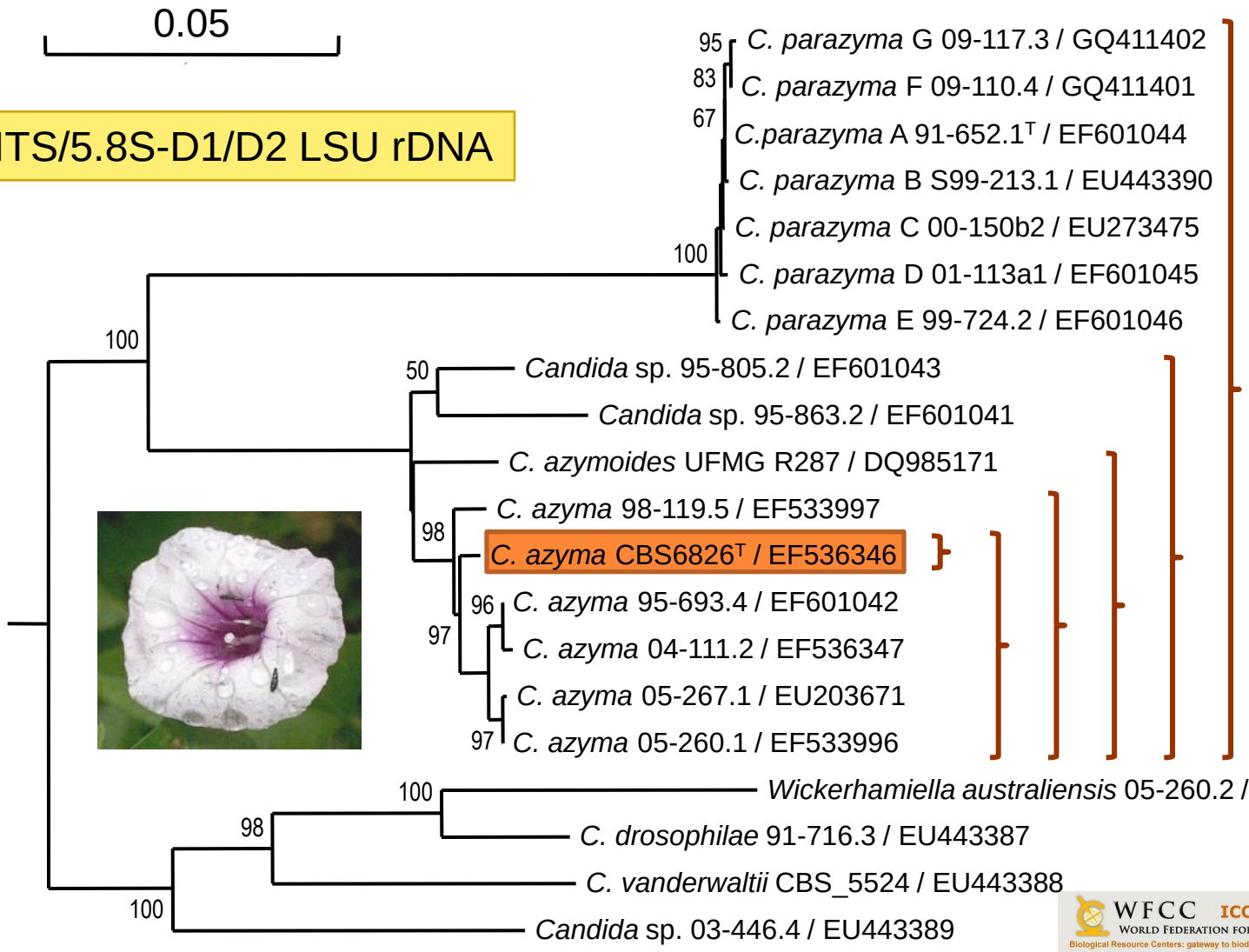


WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

CASE: *CANDIDA AZYMA*

Lachance,
Dobson,
D Wijayanayaka,
Smith
2010

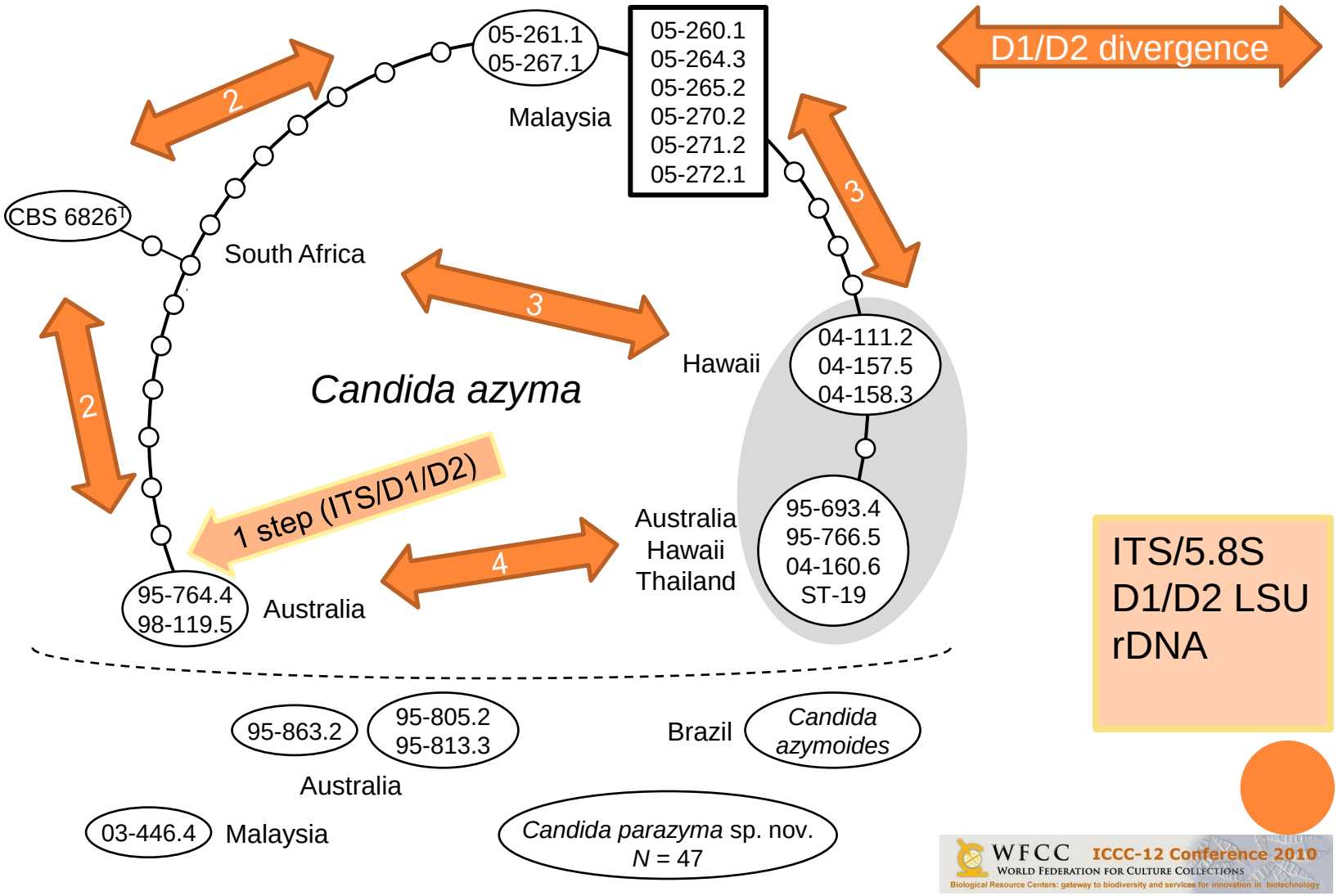
Western Biology



WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

CASE: *CANDIDA AZYMA*

Lachance,
Dobson,
D Wijayanayaka,
Smith
2010

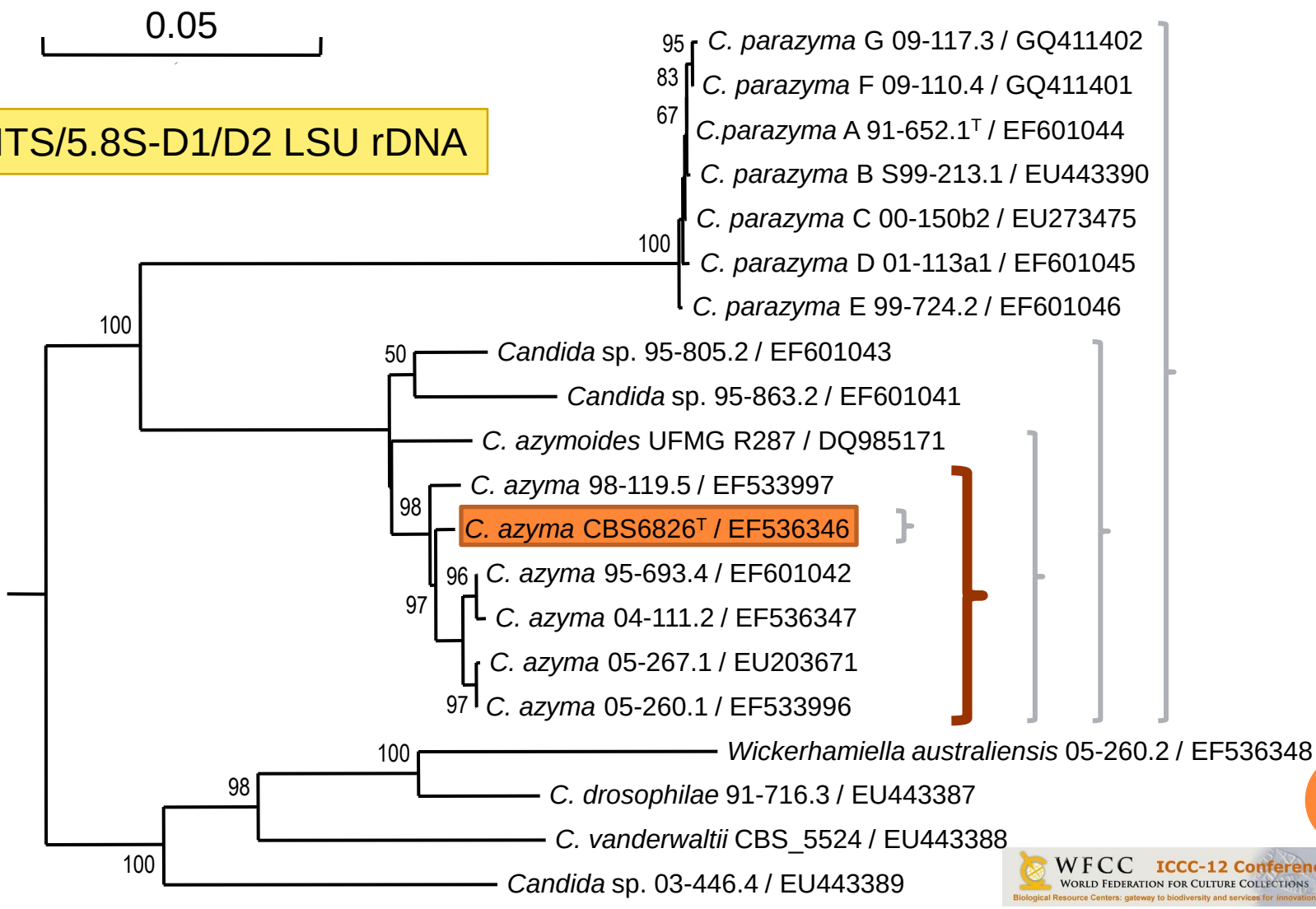


ITS/5.8S
D1/D2 LSU
rDNA

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

CASE: *CANDIDA AZYMA*

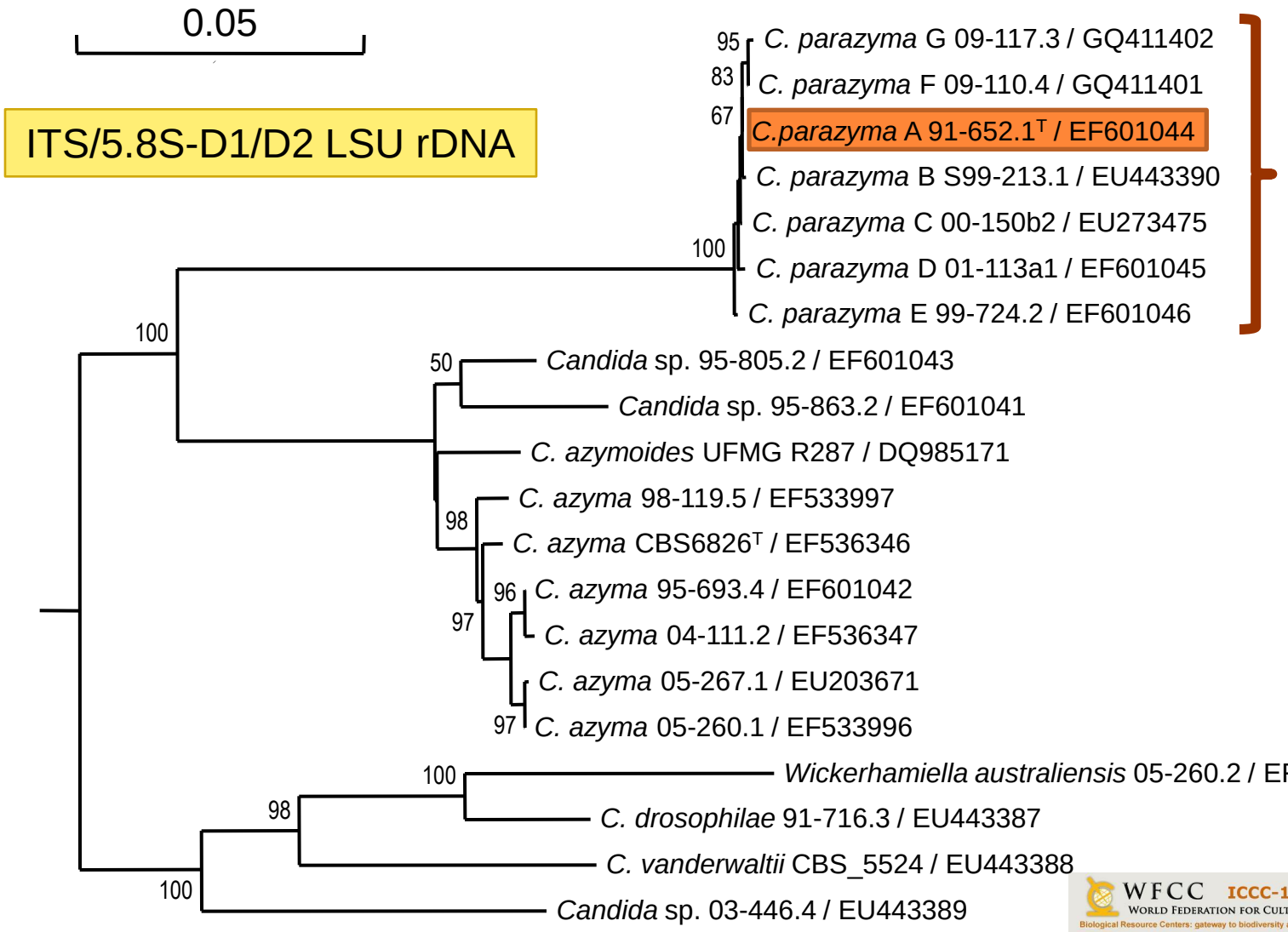
Lachance,
Dobson,
D Wijayanayaka,
Smith
2010



WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

CASE: *CANDIDA PARAZYMA*

Lachance,
Dobson,
D Wijayanayaka,
Smith
2010

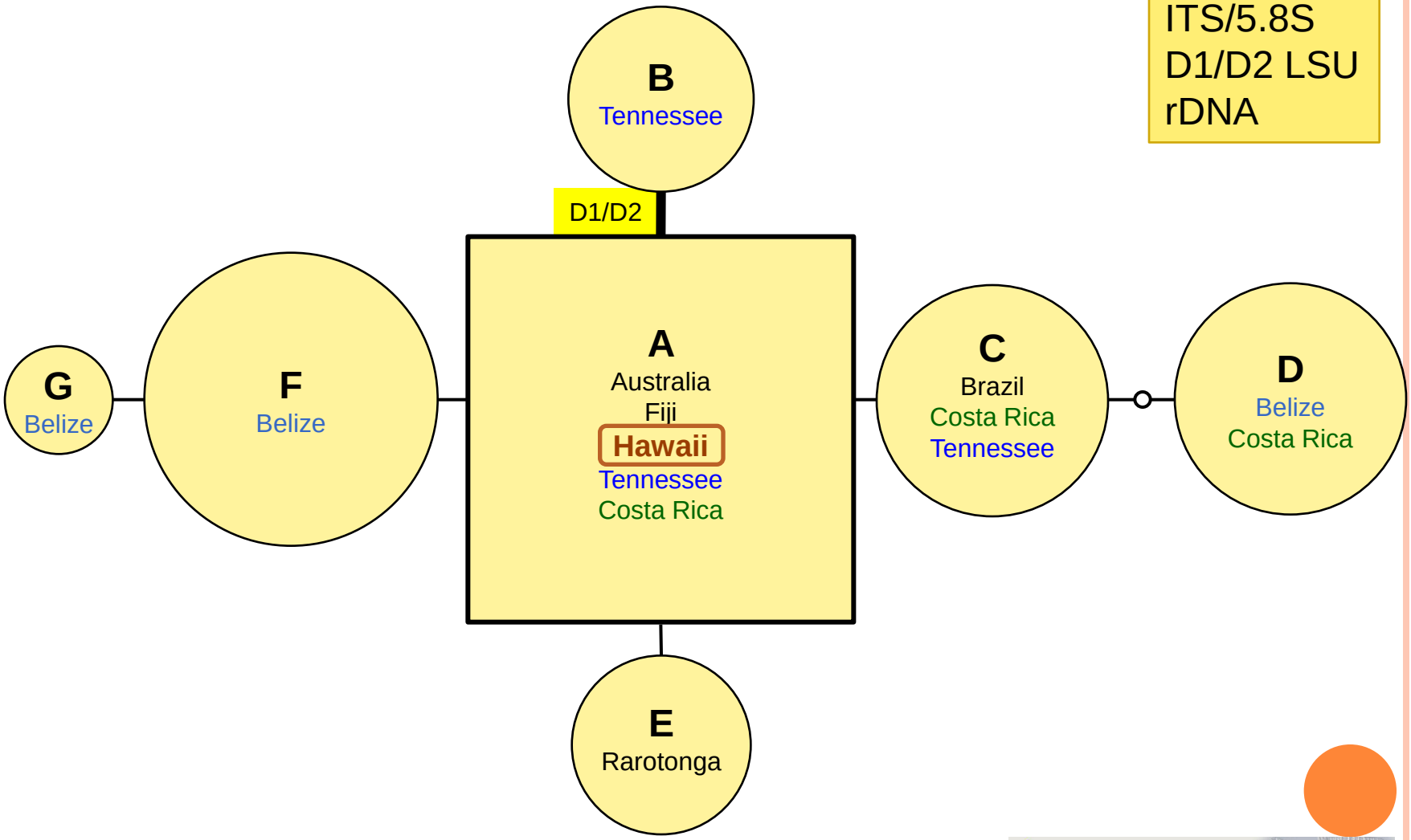


WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

CASE: *CANDIDA PARAZYMA*

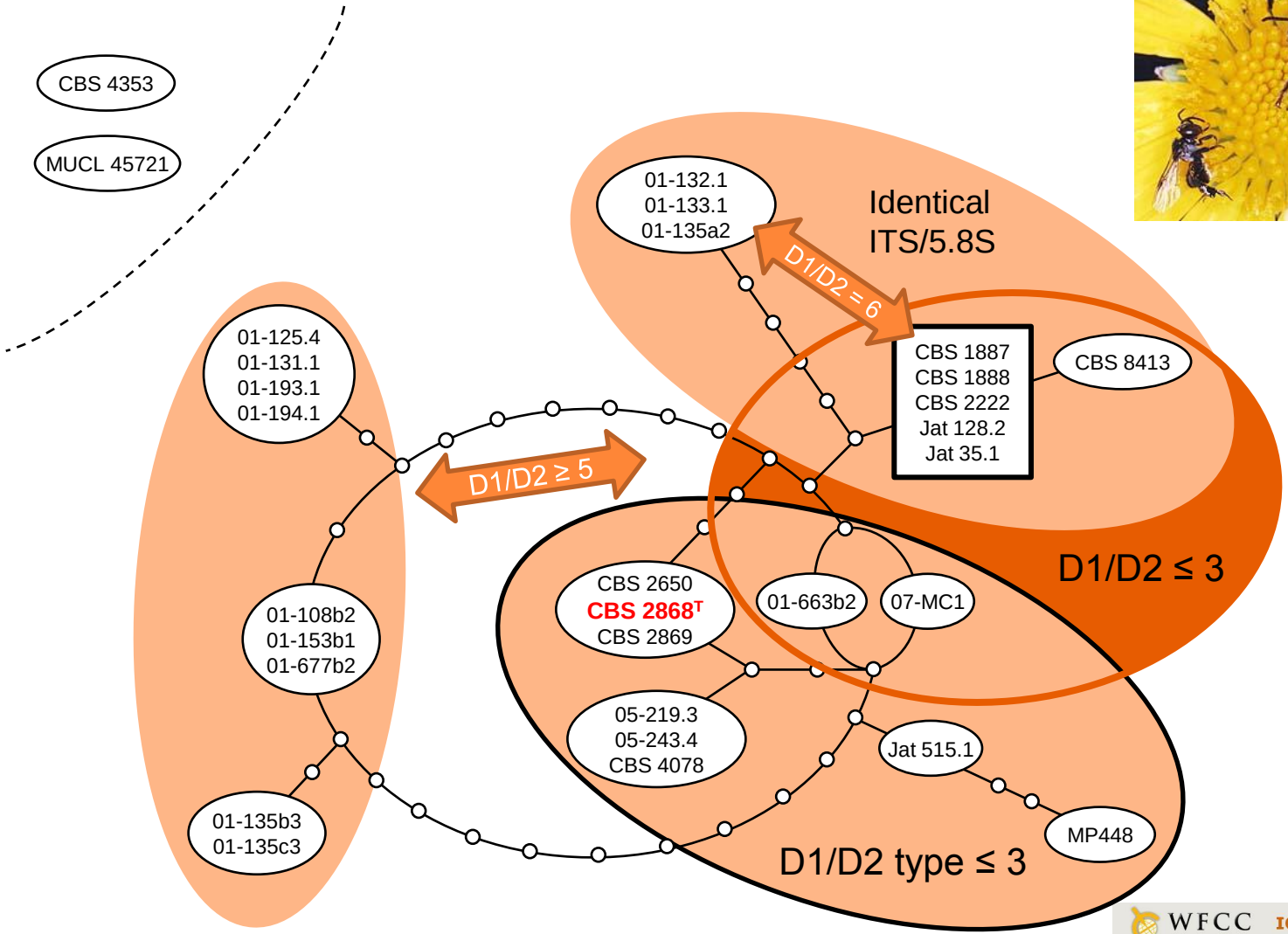
Lachance,
Dobson,
D Wijayanayaka,
Smith
2010

ITS/5.8S
D1/D2 LSU
rDNA



CASE: *CANDIDA APICOLA*

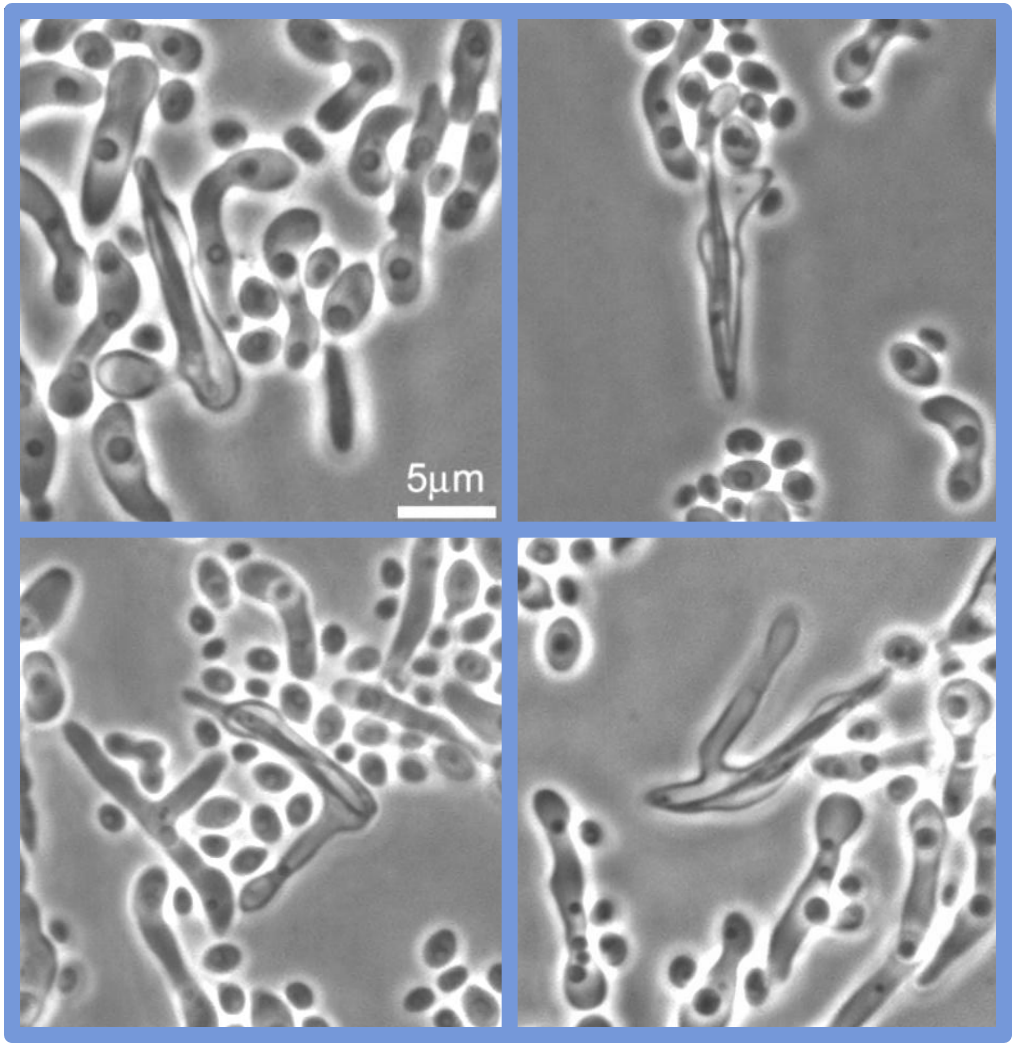
Lachance,
Dobson,
D Wijayanayaka,
Smith
2010



ITS/5.8S
D1/D2 LSU
rDNA

CASE: *METSCHNIKOWIA AGAVES*

Bundus & Lachance
unpublished

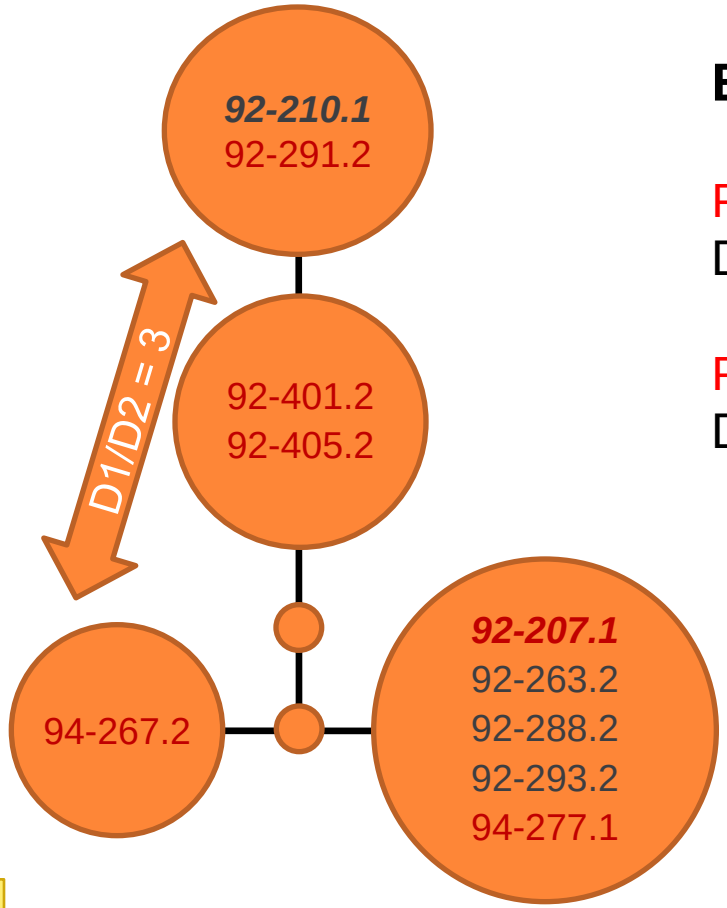


N = 10

All (6×4) compatible crosses produce asci with two ascospores

CASE: *M. AGAVES*

Bundus & Lachance
unpublished



ITS/5.8S
D1/D2 LSU
rDNA

Extrapolations:

Predicted maximum number of D1/D2 variants in species ≈ 6

Predicted maximum number of D1/D2 substitutions in species ≈ 4-5

Candida wancherniae

Nakase et al. 2009

CASE: *STARMERELLA BOMBICOLA*

T. Wijayanayaka
& Lachance
unpublished

Western Biology

ITS/5.8S
D1/D2 LSU
rDNA

All included strains can be connected to a neighbour by 3 or fewer D1/D2 substitutions

Ontario
Europe
99-213.2
EBD147
EBD176
99-202.2
CBS 9710
CBS 9711

7 D1/D2

7 D1/D2

Hawai'i
00-623c3

Europe
DBV3175 DBV 4120
DBV3176 DBV 4121
DBV3708 DBV 4122
DBV3711 DBV 4126
DBV3932 CBS 5710

00-531
00-227.2

CBS7267

00-141.1 Costa Rica

Ontario
99-543.1
99-546.1
00-401.1
00-502.1

01-123.1 Costa Rica

Ontario
99-547.2

7 D1/D2

New York
97-211.1

Identical D1/D2

97-105 97-118
97-107 97-119
97-112
97-116 **CBS 6009**

Canada
99-542.1
Ontario

05-236.1
Malaysia

CBS6366
South Africa

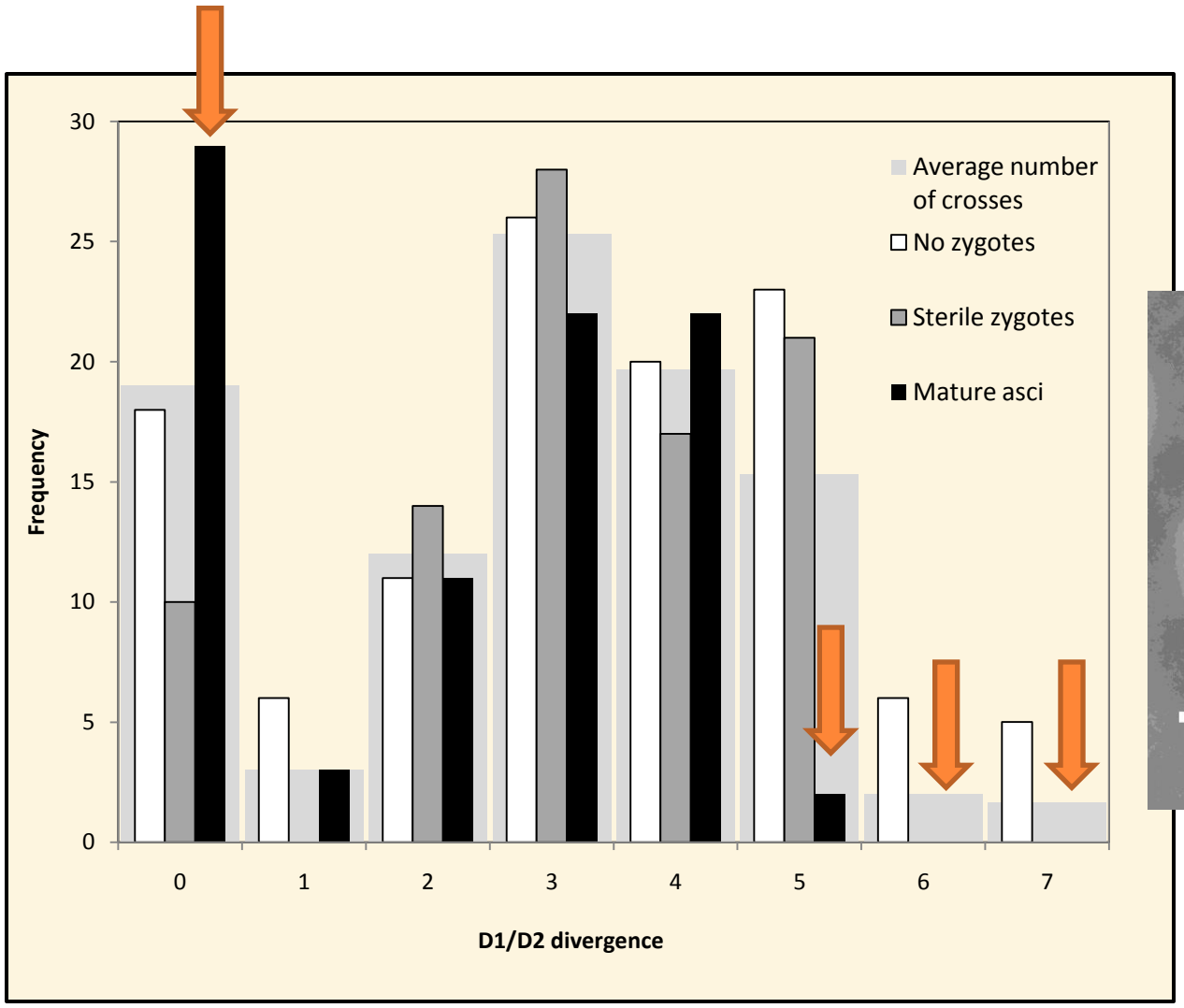
Identical D1/D2

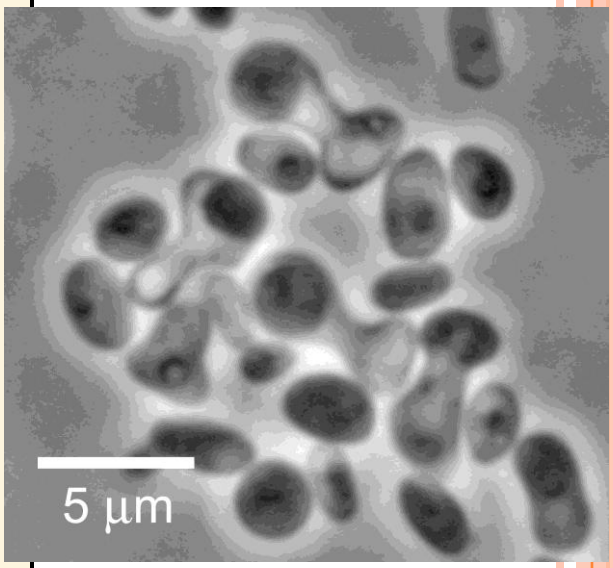
99-686.3
Fiji

WFCC ICCC-12 Conference 2010
World Federation for Culture Collections
Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

CASE: *STARMERELLA BOMBICOLA*

T. Wijayanayaka & Lachance unpublished





SUMMARY

- Barcode DNA sequencing
  - has become the standard for yeast identification
  - provides preliminary phylogenetic information
- Application to species delineation must be rooted in a sound theoretical framework
  - Adequate sampling of species is essential
  - ITS/5.8S-D1/D2 haplotype networks help in discriminating between tokogenetic (within-species) and phylogenetic (between species) relationships

ACKNOWLEDGEMENTS

- Lab work
  - Jessica Dobson
  - Dilini Wijayanayaka
  - Allison Smith
  - Joanna Bundus
  - Tishara Wijayanayaka
- Inspiration
  - Daria Koscinski
  - Clete Kurtzman
  - Carlos Rosa
  - Tom Starmer

Western Biology

NSERC CRSNG