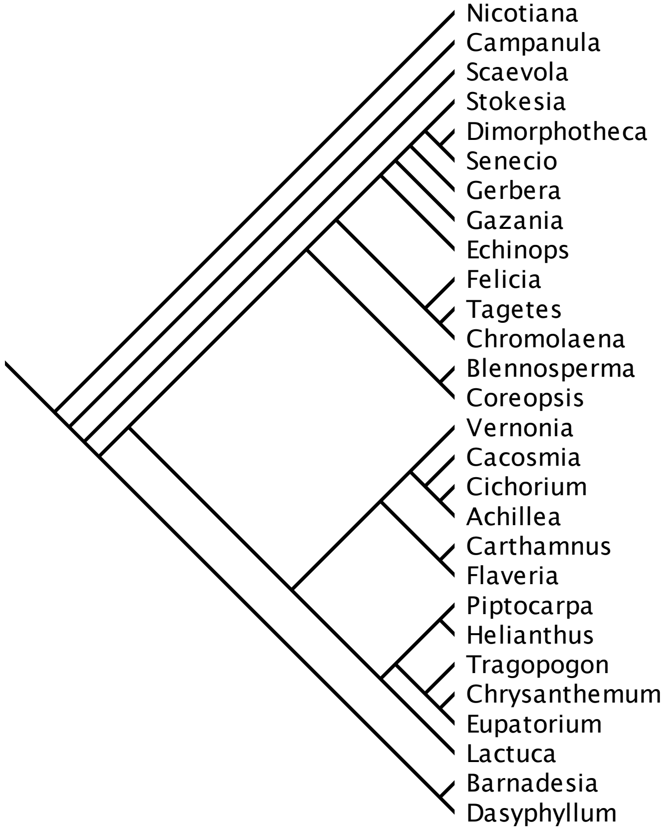
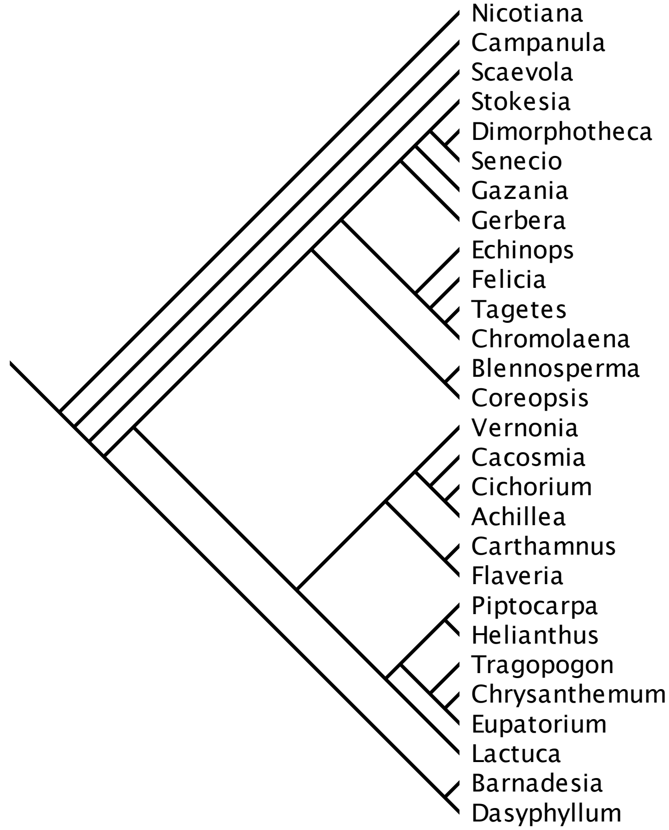
# The Complexity of the uSPR Distance

Maria Luisa Bonet
Universtidad Politecnica de Catalunya
Barcelona, Spain
bonet@lsi.upc.edu

Published in IEEE/ACM TCBB 2010
Joint work with Katherine St. John

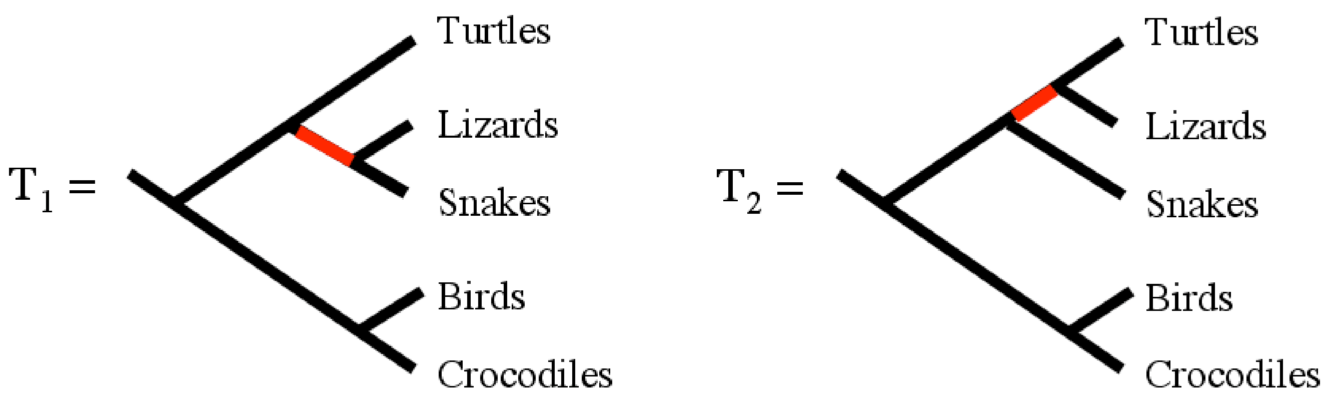
# When are two trees similar?




Phylogenies for sunflowers. Bob Jansen (UT Austin).

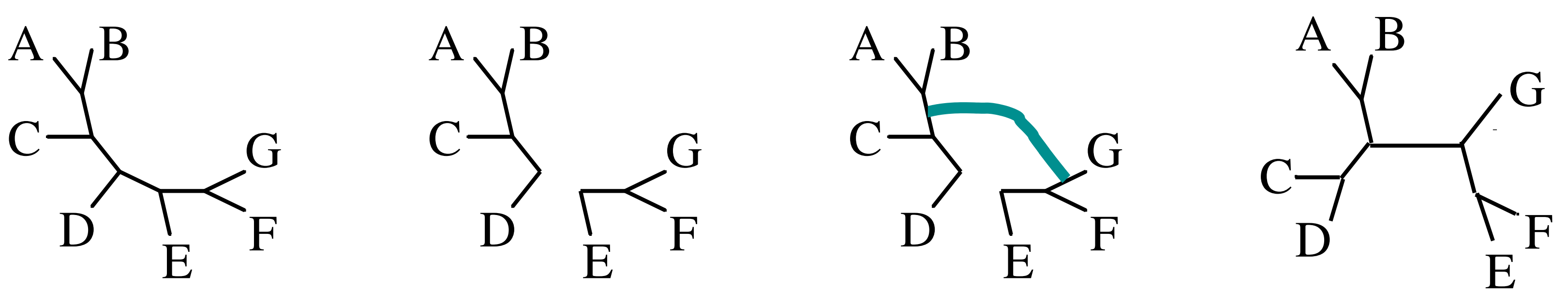
# Distances Between Trees




- Robinson-Foulds distance: # of branches that occur in only one tree.
- Calculate in $O(n)$ time using Day's Algorithm (1985).
- Extends naturally to weighted trees.

# TBR Distance

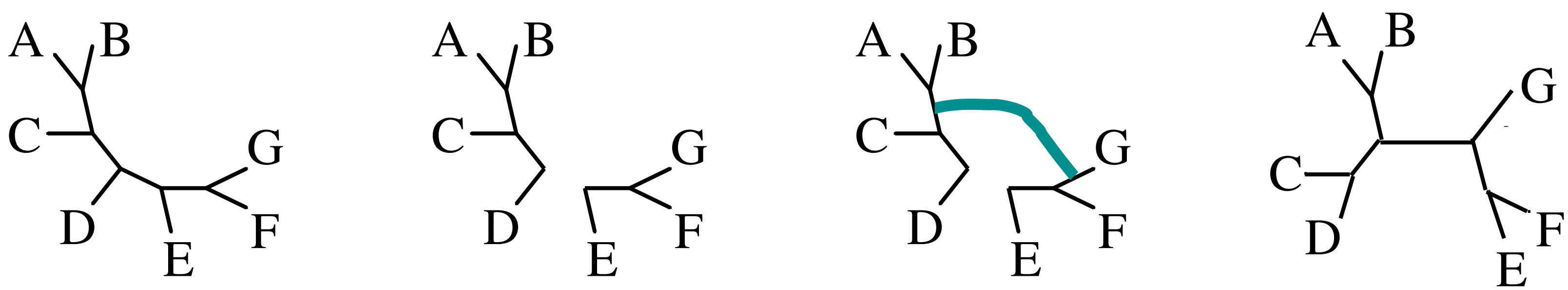
- Tree-Bisection-Reconnect (TBR) Move:




- The TBR distance between two trees is the minimal number of TBR moves needed to transform the first tree into the second tree.

# TBR Distance

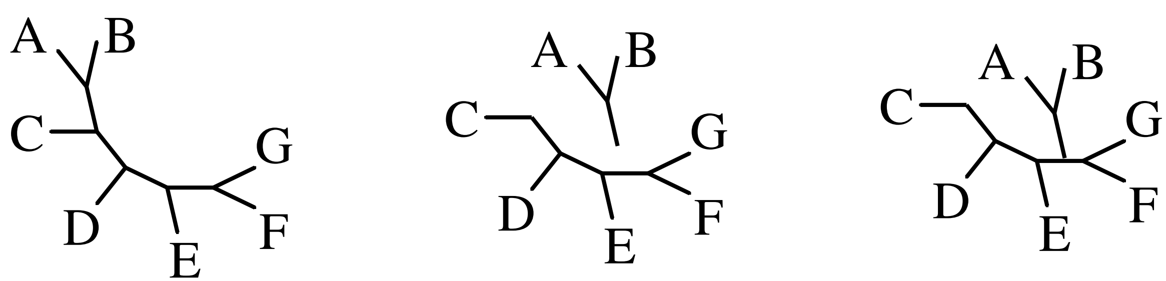
- Tree-Bisection-Reconnect (TBR):




- TBR is NP-hard. (Allen & Steel '01)

# SPR Distance

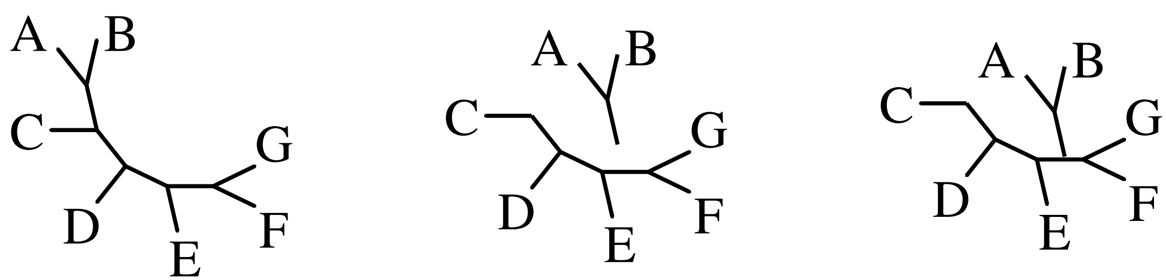
- Subtree-prune-regraft (SPR):




- The SPR distance between two trees is the minimal number of SPR moves needed to transform the first tree into the second tree.
- $d_{TBR}(T_1, T_2) \leq d_{SPR}(T_1, T_2)$

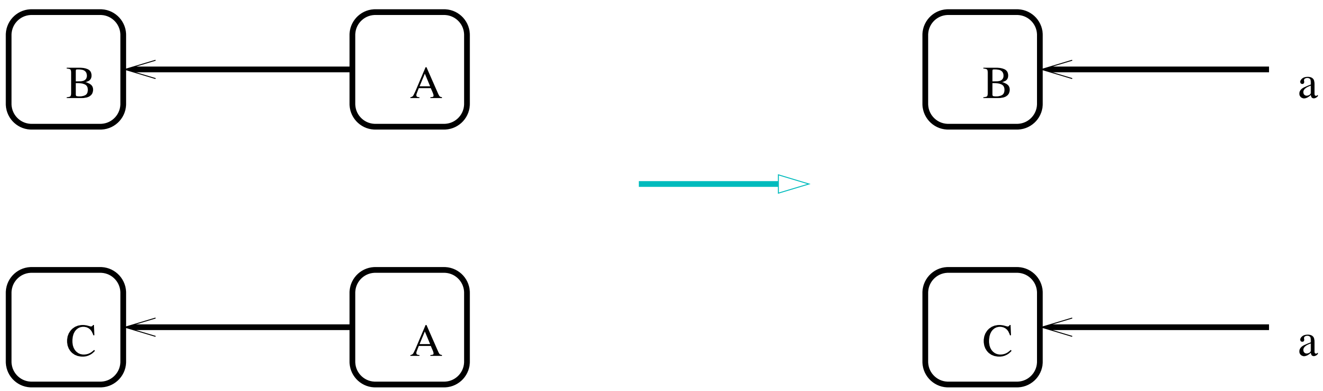
# SPR Distance

- Subtree-prune-regraft (SPR):



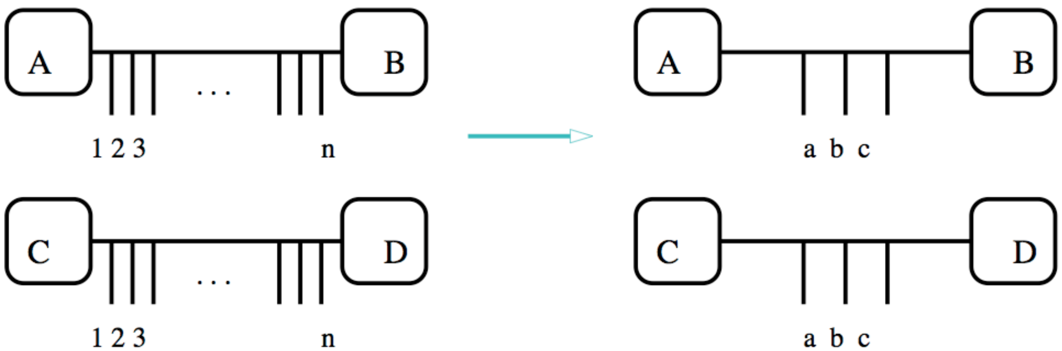

- SPR for rooted trees is NP-hard. (Bordewich & Semple '05)
- Approximation algorithm for SPR on rooted trees (Bonet, St. John, Amenta, & Mahindru '05) (Borderwich, McCartin, Semple).

# Reduction Rules: Subtree Rule




Applying Subtree Rule preserves TBR distance (Allen & Steel ‘01), SPR distance (Allen &Steel ‘01, Borderwich & Semple ‘04) and Hybrid number (Borderwich & Semple ‘07).

# Reduction Rules: Chain Rule




Applying the chain Rule preserves TBR distance (Allen & Steel '01), rSPR distance (Borderwich & Semple '04) and Hybrid number (Borderwich & Semple '07).

# $100 Problem

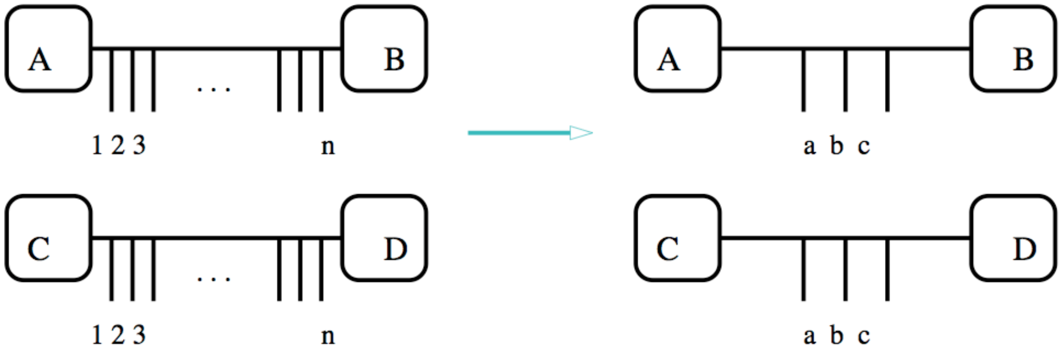
Mike Steel posed the following questions:

Does shrinking common chains preserve SPR distance?

Calculating SPR distance is fixed parameter tractable?

Calculating TBR (Allen & Steel '01), rSPR (Borderwich & Semple '04), or Hybrid Number (Borderwich & Semple '07) is fixed parameter tractable.

# Lower Bounds for Subchain Reduction




Hickey *et al.* showed that:

- Calculating SPR distance is NP-hard.
- $d_{uSPR}(T_1^n, T_2^n) - 2 \leq d_{uSPR}(T_1^3, T_2^3)$.
- We improve this to: $d_{uSPR}(T_1^n, T_2^n) - 1 \leq d_{uSPR}(T_1^3, T_2^3)$

# SPR distance is FPT

We show that uSPR is fixed parameter tractable with respect to parameter $k$:

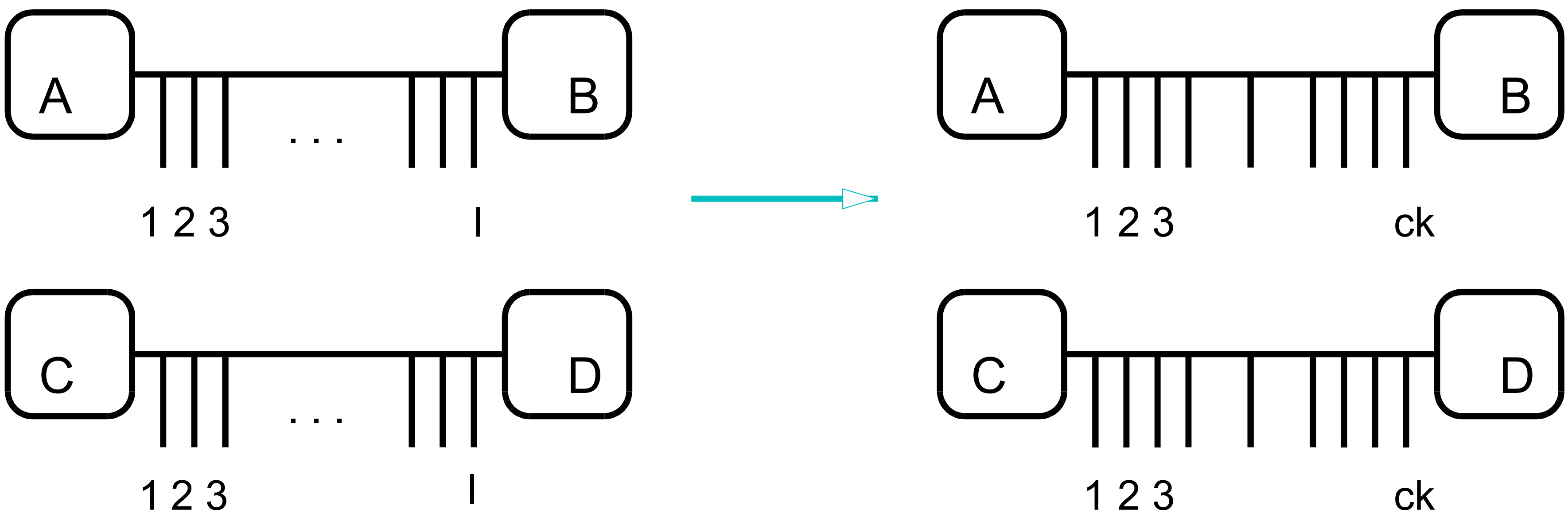
That is, for distance $k$, it can be decided in $p(n)f(k)$ time that two trees are distance $k$,

- where $p$ is a polynomial in $n$, the number of leaves in the tree, and
- $f(k)$ does not depend on $n$.

# Sketch: TBR distance is FPT

- Input: Pair of trees $(T_1, T_2)$ and distance parameter $k$.
- Obtain $(T_1', T_2')$ applying subtree and clain reduction. $d_{TBR}(T_1, T_2) = d_{TBR}(T_1', T_2')$ unknown for SPR
- $|T_1'| \leq c.d_{TBR}(T_1, T_2)$ (lemma in Allen-Steel 2001)
- if $|T_1'| > c.k$ then $d_{TBR}(T_1, T_2) > k$ and return "no".
- else check all posible $k$ TBR-moves from $T_1'$ to $T_2'$ and return "yes" or "no".
- This can be done in time $O((|T_1'|^2)^k) = O((d \cdot k)^{2k})$ Total time is $O(k^{2k} p(n))$.

# New Reduction Rule: l-Chain Rule




Lemma: Let $T_1$ and $T_2$ be X-trees with SPR distance $k$. Applying the $c \cdot k$-chain Rule preserves SPR distance.

# Size of Reduced Trees

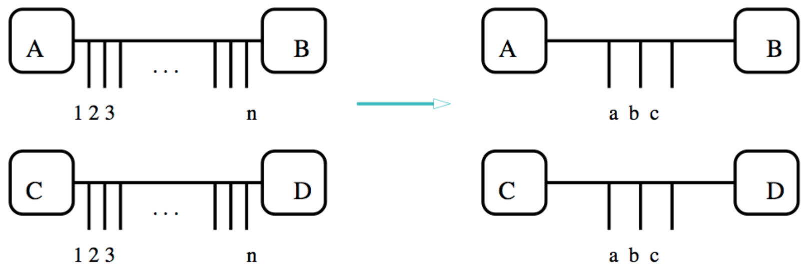
Lemma: Let $T_1$ and $T_2$ be X-trees with SPR distance $k$. Let $T_1'$ and $T_2'$ be the trees obtained from $T_1$ and $T_2$ applying subtree and $c \cdot k$-chain Rule. Then, $|T_1'| \leq dk^2$.

Proof:
if $d_{SPR}(T_1, T_2) \leq k$ then $d_{TBR}(T_1, T_2) \leq k$. Show that if $d_{TBR}(T_1, T_2) \leq k$, then $|T_1'| \leq dk^2$ as in Allen-Steel 2001 but with c.k-chains instead of 3-chains.

# Sketch: SPR distance is FPT

- Input: Pair of trees $(T_1, T_2)$ and distance parameter $k$.
- Obtain $(T_1', T_2')$ applying subtree and ck-clain reduction.
- If $d_{SPR}(T_1, T_2) \leq k$, $|T_1'| \leq d \cdot k^2$ and $d_{SPR}(T_1, T_2) = d_{SPR}(T_1', T_2')$
- if $|T_1'| > d \cdot k^2$ then $d_{SPR}(T_1, T_2) > k$ and return "no".
- else check all posible $k$ SPR-moves from $T_1'$ to $T_2'$ and return "yes" or "no".
- This can be done in time $O((|T_1'|^2)^k) = O((d \cdot k^2)^{2k})$ Total time is $O(k^{4k} p(n))$.

# Idea for Improving Lower Bound




$$d_{uSPR}(T_1^n, T_2^n) - 1 \leq d_{uSPR}(T_1^3, T_2^3)$$

- Idea inspired from Hickey *et al*: treat common chains as subtrees to get "$-2$" bound.
- Go carefully by cases on a minimal set of moves to improve the bound to "$-1$".

# Idea for Improving Lower Bound

Let $T_1$ and $T_2$ X-trees with common chain $1 \cdots l$ and SPR dist $k$.

We show $d_{SPR}(T_1^3, T_2^3) \geq k - 1$ by cases.

Case sketch: the minimal number of moves breaks 2 or 3 pendant edges from the 3-chain.