Phylogeny of carnivorous plants

comparison of basic algorithms on real data

K. Tucek

January 25, 2017

Goals

This project was made as a part of fundamental course of bio-algorithms. The results are by no means to be interpretted as scientically relevant! Our aim is to:

- Try basic DNA-sequence similarity algorithms on a real problem and real data.
- To see how these simple approaches behave compared to actual scientifical results.
- To compare the behaviour of some basic and sometimes naive algorithms.
- To learn something about carnivorous plants.

Topic choice.

We have chosen to try the construction of phylogeny trees of some representants of carnivorous plants. The class of carnivorous plants was picked quite arbitrarily.
This problem seems interesting since the carnivorous properties of these plants are somehow exotic and structurally interesting, since hypotheses about structural relationships being involved in phylogeny of carnivorous plants come as natural.

Classification of Carnivorous Plants

We have picked one representants from every genus of carnivorous plants. We have picked multiple representants from the Drosera genera. The following slides sum up the order/family/genera hierarchy (nonbranching families are ommited):

- Caryophyllales
- Ericales
- Lamiales
- Oxalidales
- Poales

Classification of Carnivorous Plants

- Caryophyllales
  - Aldrovanda (1 sp, Everywhere) (rare underwater plant)
  - Droseraceae
    - Dionaea (1 sp, NAm) (pitcher flytrap)
    - Drosera (233+ sp, Ev) (sticky flypaper)
    - Drosophyllum (1 sp, Spain) (suffocating flypaper)
  - Nepenthes (131+ sp, SEAs) (classical mawed pitcher plant)
  - Triphyophyllum (1 sp, Africa) (climbing vine flypaper)
- Ericales
  - Sarraceniaceae
    - Darlingtonia (1 sp, NAm) (pitfall pitcher)
    - Heliamphora (25+ sp, SAm) (heli-amphora)
    - Sarracenia (15 sp, NAm) (trumpet pitchers)
  - Roridula (2 sp, SAf) (non-digestive resin (!) flypaper)

Classification of Carnivorous Plants

- Lamiales
  - Byblis (8+ sp, Au) (carnivorous via hosted insects)
  - Philcoxia (3 sp, SAm) (sand-underground flypaper)
  - Lentibulariaceae
    - Genlisea (27+ sp, SAm, SAf) (two types of pitfall traps)
    - Pinguicula (102+ sp, Ev) (flypaper with efficient diegestion)
    - Utricularia (235+ sp, Ev) (underwater bladder traps)
- Oxalidales
  - Cephalotus (1 sp, Au) (another pitcher with actual 'teeth')
- Poales
  - Brocchinia (2 sp, SAm) (minimalistic pitfall pitcher)
  - Catopsis (1 sp, Am) (leafed pitfall trap)

Trap Types

Basic trap types are the following:

- snapping traps - such as the fly-trap
- sticky 'flypaper' traps - plants with leaves covered by a sticky extract
- pitcher traps - traps relying on the insect falling into a liquid-filled pitch and drowning inside
- other pitfall traps - other traps which rely on insect's getting in and not being able to get out
- special underwater traps

Methods overview

- We compare the organisms by the so-called 'rbcL' plastid gene (ribulose-1,5-biposphate carboxylase/oxygenase), which is one of substantive photosynthesis genes. (As used in paper by V. A. Albert, S. E. Williams and M. W. Chase)
- Distance matrix is constructed using various distance functions.
- Phylogeny tree is computed by the biopython's upgma algorithm.

Compared Algorithms

- GenCompress compression distance.
- Simple levenstein dinstance using gap-aware rating.
- Uni/Bi - directional histogram.
- Uni/Bi - directional Dotplot area with 2/3 filtering window ratio.

Compared Algorithms

Bi/Uni directionality

- In order to take into account the possibility of complement-stand crossovers, we implement the histogram and dotplot algorithms in bidirectional variants.
- The bi-directional versions consider complement's patterns as if these were found in the forward algorithm.
- This consideration is mostly theoretical since We assume that crossovers with the complement strands are not of high interes inside of gene sequences. Also, this penalty-less rating in reverse strand is naive.

Compared Algorithms

Dotplot vs histogram algorithms

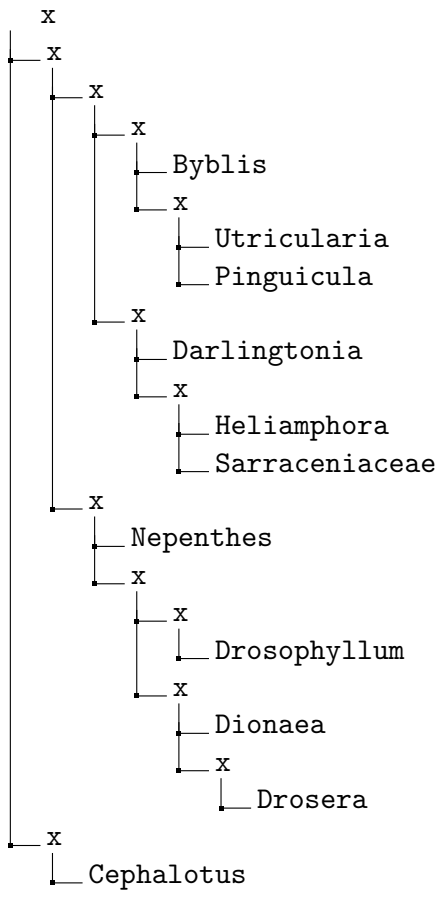
These two algorithms perform basically the same thing. However, unlike simple histogram algorithm, the dotplot surface algorithm may take into account:

- Locality of patterns.
- Mutations.
- Longer window, wince the results do not need to be explicitly stored.

On the other hand, the dotplot-surface takes reccuring patterns into account $O(n^2)$ times unlike the simple histogram algorithm. Also, dotplot algorithm is much more prone to badly tuned parameters.

Expected results

expected:



GenCompress distance

compression:

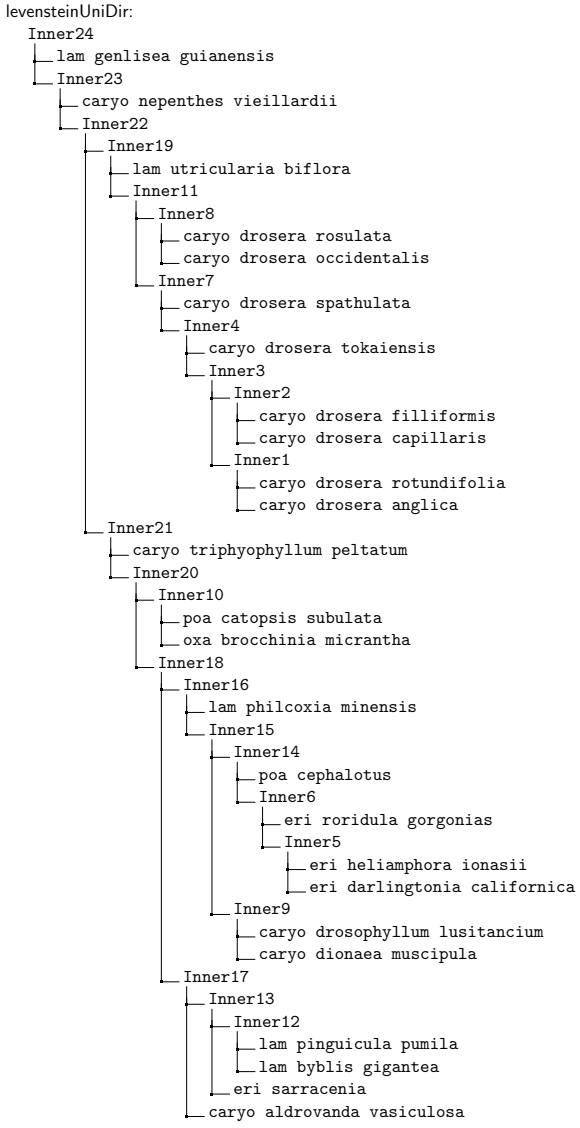
- Inner24
  - caryo nepenthes vieillardii
  - Inner23
    - Inner9
      - poa catopsis subulata
      - oxa brocchinia micrantha
    - Inner22
      - Inner21
        - Inner19
          - Inner15
            - lam utricularia biflora
            - lam genlisea guianensis
          - Inner13
            - lam pinguicula pumila
            - Inner12
              - lam philcoxia minensis
              - lam byblis gigantea
        - Inner16
          - poa cephalotus
          - Inner10
            - Inner7
              - eri sarracenia
              - Inner4
                - eri roridula gorgonias
                - eri heliamphora ionasii
            - eri darlingtonia californica
      - Inner20
        - caryo triphyophyllum peltatum
        - Inner18
          - Inner11
            - Inner6
              - caryo drosera rosulata
              - caryo drosera occidentalis
            - Inner5
              - caryo drosera tokaiensis
              - Inner3
                - Inner2
                  - caryo drosera filliformis
                  - caryo drosera capillaris
                - Inner1
                  - caryo drosera rotundifolia
                  - caryo drosera anglica
          - Inner17
            - Inner14
              - caryo drosera spathulata
              - Inner8
                - caryo drosophyllum lusitancium
                - caryo dionaea muscipula
            - caryo aldrovanda vasiculosa

expected:

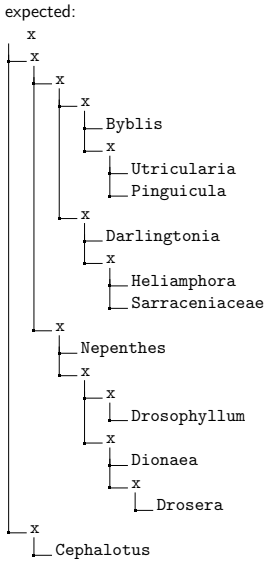
- x
  - x
    - x
      - x
        - Byblis
        - x
          - Utricularia
          - Pinguicula
      - x
        - Darlingtonia
        - x
          - Heliamphora
          - Sarraceniaceae
    - x
      - Nepenthes
      - x
        - x
          - Drosophyllum
        - x
          - Dionaea
          - x
            - Drosera
  - x
    - Cephalotus

Levenstein distance

levensteinUniDir:



expected:



Unidirectional histogram

histogramUniDir:

- Inner24
  - caryo nepenthes vieillardii
  - Inner23
    - lam genlisea guianensis
    - Inner22
      - Inner8
        - poa catopsis subulata
        - oxa brocchinia micrantha
      - Inner21
        - Inner20
          - Inner18
            - lam utricularia biflora
            - Inner13
              - lam pinguicula pumila
              - Inner11
                - lam philcoxia minensis
                - lam byblis gigantea
          - Inner16
            - poa cephalotus
            - Inner9
              - Inner7
                - eri sarracenia
                - eri heliamphora ionasii
              - Inner6
                - eri roridula gorgonias
                - eri darlingtonia californica
        - Inner19
          - Inner14
            - caryo triphyophyllum peltatum
            - caryo drosophyllum lusitancium
          - Inner17
            - Inner10
              - caryo drosera rosulata
              - caryo drosera occidentalis
            - Inner15
              - caryo dionaea muscipula
              - Inner12
                - Inner5
                  - Inner1
                    - caryo drosera tokaiensis
                    - caryo drosera spathulata
                  - Inner4
                    - Inner3
                      - caryo drosera filliformis
                      - caryo drosera capillaris
                    - Inner2
                      - caryo drosera rotundifolia
                      - caryo drosera anglica
                - caryo aldrovanda vasiculosa

expected:

- x
  - x
    - x
      - x
        - Byblis
        - x
          - Utricularia
          - Pinguicula
      - x
        - Darlingtonia
        - x
          - Heliamphora
          - Sarraceniaceae
    - x
      - Nepenthes
      - x
        - x
          - Drosophyllum
        - x
          - Dionaea
          - x
            - Drosera
  - x
    - Cephalotus

Unidirectional dotplot surface area

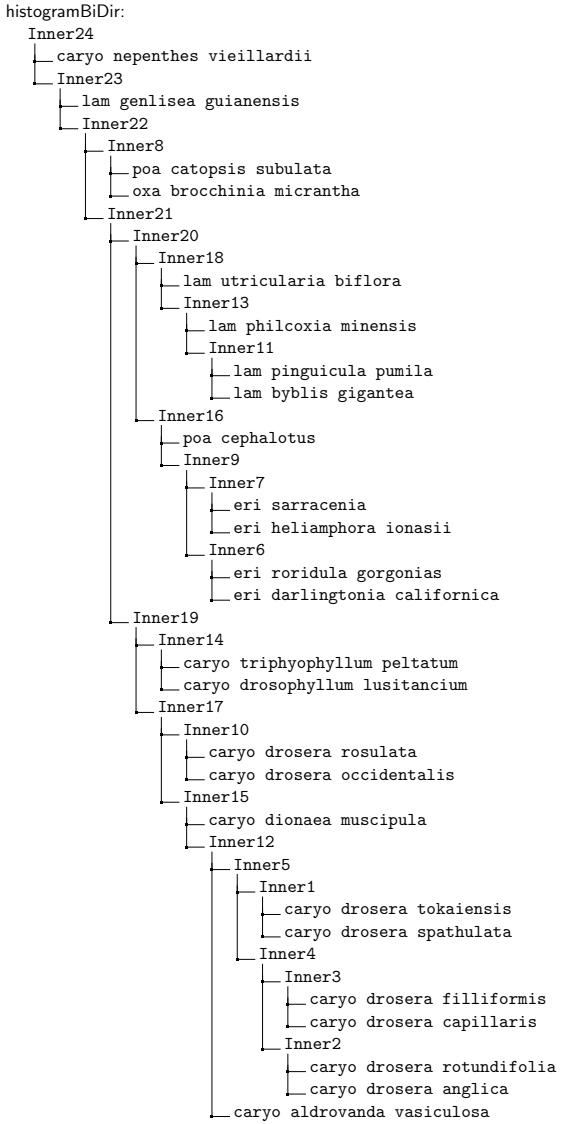
dotplotUniDir:

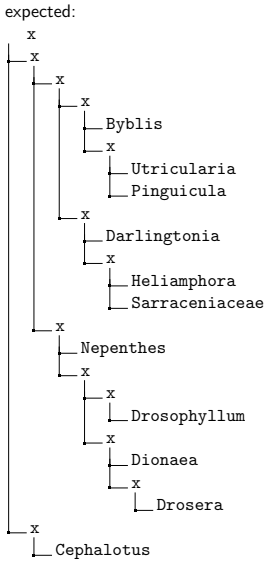
- Inner24
  - caryo nepenthes vieillardii
  - Inner23
    - Inner11
      - poa catopsis subulata
      - oxa brocchinia micrantha
    - Inner22
      - Inner21
        - Inner16
          - lam philcoxia minensis
          - Inner14
            - lam pinguicula pumila
            - Inner13
              - Inner8
                - lam utricularia biflora
                - lam genlisea guianensis
              - lam byblis gigantea
        - Inner19
          - poa cephalotus
          - Inner10
            - Inner6
              - eri sarracenia
              - eri heliamphora ionasii
            - Inner9
              - eri roridula gorgonias
              - eri darlingtonia californica
      - Inner20
        - Inner18
          - caryo triphyophyllum peltatum
          - caryo drosophyllum lusitancium
        - Inner17
          - caryo dionaea muscipula
          - Inner15
            - Inner12
              - Inner7
                - caryo drosera rosulata
                - caryo drosera occidentalis
              - Inner5
                - caryo drosera spathulata
                - Inner4
                  - caryo drosera filliformis
                  - Inner3
                    - Inner2
                      - caryo drosera tokaiensis
                      - caryo drosera capillaris
                    - Inner1
                      - caryo drosera rotundifolia
                      - caryo drosera anglica
            - caryo aldrovanda vasiculosa

expected:

- x
  - x
    - x
      - x
        - Byblis
        - x
          - Utricularia
          - Pinguicula
      - x
        - Darlingtonia
        - x
          - Heliamphora
          - Sarraceniaceae
    - x
      - Nepenthes
      - x
        - x
          - Drosophyllum
        - x
          - Dionaea
          - x
            - Drosera
  - x
    - Cephalotus

Bidirectional histogram





GenCompress distance

caryo aldrovanda vasiculosa	0																												
caryo dionaea muscipula	0.76	0																											
caryo drosera anglica	0.76	0.80	0																										
caryo drosera capillaris	0.77	0.82	0.21	0																									
caryo drosera filliformis	0.78	0.81	0.22	0.19	0																								
caryo drosera occidentalis	0.79	0.80	0.60	0.64	0.63	0																							
caryo drosera rosulata	0.79	0.81	0.63	0.65	0.66	0.51	0																						
caryo drosera rotundifolia	0.77	0.81	0.17	0.24	0.26	0.60	0.63	0																					
caryo drosera spathulata	0.78	0.72	0.51	0.52	0.54	0.78	0.77	0.51	0																				
caryo drosera tokaiensis	0.77	0.81	0.49	0.48	0.50	0.75	0.78	0.49	1.0	0																			
caryo drosophyllum lusitancium	0.80	0.61	0.81	0.82	0.83	0.82	0.83	0.82	0.73	0.82	0																		
caryo nepenthes vieillardii	0.94	0.98	0.98	0.98	1.00	0.93	0.92	0.98	1.00	1.00	0.97	0																	
caryo triphyophyllum peltatum	0.81	0.82	0.82	0.83	0.83	0.84	0.84	0.82	0.85	0.86	0.73	0.98	0																
eri darlingtonia californica	0.85	0.85	0.91	0.91	0.92	0.90	0.90	0.91	0.90	0.90	0.80	0.95	0.83	0															
eri heliamphora ionasii	0.84	0.90	0.93	0.93	0.93	0.91	0.89	0.93	0.92	0.92	0.87	0.95	0.83	0.60	0														
eri roridula gorgonias	0.82	0.88	0.92	0.92	0.92	0.89	0.90	0.92	0.90	0.90	0.84	0.93	0.82	0.61	0.37	0													
eri sarracenia	0.84	0.90	0.93	0.92	0.93	0.89	0.89	0.93	0.92	0.91	0.86	0.95	0.84	0.66	0.53	0.62	0												
lam byblis gigantea	0.85	0.84	0.90	0.92	0.92	0.83	0.86	0.91	0.83	0.91	0.81	0.90	0.87	0.80	0.78	0.78	0.81	0											
lam genlisea guianensis	0.84	0.94	0.96	0.96	0.96	0.95	0.94	0.96	0.97	0.96	0.95	0.96	0.94	0.85	0.86	0.86	0.84	0.85	0										
lam philcoxia minensis	0.85	0.92	0.93	0.92	0.93	0.89	0.93	0.93	0.93	0.92	0.91	0.95	0.90	0.83	0.79	0.78	0.79	0.71	0.83	0									
lam pinguicula pumila	0.85	0.92	0.87	0.85	0.85	0.83	0.85	0.88	0.93	0.91	0.87	0.96	0.89	0.82	0.75	0.76	0.74	0.73	0.79	0.72	0								
lam utricularia biflora	0.84	0.94	0.95	0.94	0.94	0.93	0.95	0.95	0.95	0.95	0.91	0.97	0.90	0.79	0.82	0.81	0.83	0.79	0.76	0.80	0.75	0							
oxa brocchinia micrantha	0.91	0.95	0.96	0.97	0.97	0.95	0.95	0.97	0.97	0.96	0.94	0.93	0.93	0.88	0.83	0.84	0.83	0.90	0.88	0.86	0.86	0.90	0						
poa catopsis subulata	0.98	0.98	0.96	0.94	0.97	0.99	0.99	0.97	0.97	0.97	0.96	0.98	0.92	0.91	0.91	0.92	0.90	0.95	0.96	0.94	0.94	0.95	0.62	0					
poa cephalotus	0.87	0.84	0.92	0.91	0.92	0.91	0.91	0.92	0.92	0.90	0.80	0.98	0.83	0.76	0.78	0.78	0.78	0.85	0.91	0.87	0.83	0.86	0.90	0.93	0				
x car	car	car	car	car	car	car	car	car	car	car	car	car	eri	eri	eri	eri	lam	lam	lam	lam	lam	oxa	poa	poa					

Unidirectional histogram

caryo aldrovanda vasiculosa	0																												
caryo dionaea muscipula	0.32	0																											
caryo drosera anglica	0.31	0.38	0																										
caryo drosera capillaris	0.33	0.39	0.05	0																									
caryo drosera filliformis	0.33	0.39	0.05	0.03	0																								
caryo drosera occidentalis	0.36	0.39	0.31	0.34	0.34	0																							
caryo drosera rosulata	0.40	0.40	0.35	0.36	0.37	0.27	0																						
caryo drosera rotundifolia	0.31	0.38	0.01	0.07	0.07	0.32	0.35	0																					
caryo drosera spathulata	0.31	0.37	0.07	0.08	0.10	0.33	0.35	0.07	0																				
caryo drosera tokaiensis	0.32	0.38	0.05	0.06	0.08	0.32	0.34	0.06	0.01	0																			
caryo drosophyllum lusitancium	0.35	0.38	0.39	0.40	0.40	0.40	0.43	0.39	0.39	0.39	0																		
caryo nepenthes vieillardii	0.85	0.86	0.87	0.87	0.87	0.88	0.87	0.87	0.86	0.87	0.84	0																	
caryo triphyophyllum peltatum	0.41	0.43	0.46	0.47	0.48	0.46	0.47	0.47	0.46	0.46	0.34	0.85	0																
eri darlingtonia californica	0.50	0.47	0.50	0.49	0.50	0.49	0.50	0.50	0.48	0.48	0.41	0.83	0.42	0															
eri heliamphora ionasii	0.49	0.46	0.50	0.49	0.50	0.51	0.49	0.50	0.48	0.48	0.42	0.84	0.43	0.12	0														
eri roridula gorgonias	0.49	0.45	0.49	0.48	0.50	0.49	0.50	0.49	0.47	0.48	0.41	0.82	0.44	0.12	0.15	0													
eri sarracenia	0.50	0.48	0.50	0.49	0.50	0.49	0.47	0.50	0.49	0.48	0.45	0.86	0.46	0.19	0.12	0.20	0												
lam byblis gigantea	0.46	0.49	0.48	0.47	0.46	0.48	0.51	0.48	0.46	0.47	0.44	0.85	0.46	0.37	0.37	0.36	0.40	0											
lam genlisea guianensis	0.70	0.71	0.68	0.68	0.68	0.67	0.69	0.68	0.69	0.68	0.68	0.85	0.68	0.65	0.66	0.65	0.68	0.57	0										
lam philcoxia minensis	0.51	0.51	0.53	0.52	0.52	0.52	0.56	0.53	0.51	0.53	0.49	0.84	0.52	0.42	0.43	0.40	0.44	0.29	0.62	0									
lam pinguicula pumila	0.53	0.52	0.53	0.51	0.52	0.52	0.52	0.53	0.52	0.51	0.46	0.88	0.49	0.39	0.38	0.38	0.40	0.30	0.60	0.37	0								
lam utricularia biflora	0.52	0.57	0.54	0.52	0.52	0.53	0.56	0.55	0.53	0.54	0.50	0.84	0.51	0.44	0.46	0.46	0.48	0.34	0.58	0.44	0.39	0							
oxa brocchinia micrantha	0.60	0.57	0.59	0.59	0.59	0.60	0.61	0.59	0.58	0.58	0.56	0.82	0.57	0.47	0.48	0.49	0.50	0.53	0.69	0.54	0.53	0.56	0						
poa catopsis subulata	0.60	0.57	0.59	0.58	0.58	0.60	0.61	0.59	0.58	0.58	0.56	0.83	0.57	0.48	0.49	0.49	0.51	0.51	0.69	0.53	0.52	0.56	0.12	0					
poa cephalotus	0.49	0.47	0.51	0.50	0.51	0.51	0.53	0.51	0.49	0.49	0.44	0.86	0.45	0.36	0.36	0.35	0.38	0.43	0.70	0.46	0.43	0.50	0.52	0.51	0				
x car	car	car	car	car	car	car	car	car	car	car	car	car	eri	eri	eri	eri	lam	lam	lam	lam	lam	oxa	poa	poa					

Unidirectional dotplot surface area

caryo aldrovanda vasiculosa	0																														
caryo dionaea muscipula	0.99	0																													
caryo drosera anglica	0.99	0.99	0																												
caryo drosera capillaris	0.99	0.99	0.99	0																											
caryo drosera filliformis	0.99	0.99	0.99	0.99	0																										
caryo drosera occidentalis	0.99	0.99	0.99	0.99	0.99	0																									
caryo drosera rosulata	0.99	0.99	0.99	0.99	0.99	0.99	0																								
caryo drosera rotundifolia	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0																							
caryo drosera spathulata	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0																						
caryo drosera tokaiensis	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0																					
caryo drosophyllum lusitancium	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0																				
caryo nepenthes vieillardii	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0																			
caryo triphyophyllum peltatum	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0																		
eri darlingtonia californica	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0																	
eri heliamphora ionasii	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0																
eri roridula gorgonias	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0															
eri sarracenia	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0														
lam byblis gigantea	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0													
lam genlisea guianensis	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0												
lam philcoxia minensis	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0											
lam pinguicula pumila	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0										
lam utricularia biflora	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0									
oxa brocchinia micrantha	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0								
poa catopsis subulata	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0							
poa cephalotus	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0						
x car	car	car	car	car	car	car	car	car	car	car	car	car	eri	eri	eri	eri	lam	lam	lam	lam	lam	oxa	poa	poa							

Levenstein distance

caryo aldrovanda vasiculosa	0																											
caryo dionaea muscipula	0.09	0																										
caryo drosera anglica	0.14	0.19	0																									
caryo drosera capillaris	0.15	0.19	0.00	0																								
caryo drosera filliformis	0.15	0.19	0.00	0.00	0																							
caryo drosera occidentalis	0.15	0.19	0.04	0.05	0.05	0																						
caryo drosera rosulata	0.16	0.19	0.05	0.06	0.06	0.04	0																					
caryo drosera rotundifolia	0.14	0.19	0.00	0.00	0.01	0.04	0.05	0																				
caryo drosera spathulata	0.12	0.16	0.04	0.04	0.04	0.08	0.08	0.04	0																			
caryo drosera tokaiensis	0.15	0.19	0.01	0.01	0.01	0.05	0.06	0.01	0.04	0																		
caryo drosophyllum lusitancium	0.10	0.06	0.19	0.19	0.19	0.19	0.19	0.19	0.16	0.19	0																	
caryo nepenthes vieillardii	0.52	0.53	0.49	0.49	0.49	0.50	0.50	0.5	0.50	0.49	0.53	0																
caryo triphyophyllum peltatum	0.19	0.19	0.27	0.27	0.27	0.27	0.27	0.27	0.25	0.27	0.17	0.56	0															
eri darlingtonia californica	0.13	0.09	0.22	0.22	0.22	0.22	0.22	0.22	0.19	0.22	0.07	0.53	0.19	0														
eri heliamphora ionasii	0.13	0.09	0.22	0.22	0.22	0.22	0.22	0.22	0.19	0.22	0.08	0.53	0.19	0.02	0													
eri roridula gorgonias	0.13	0.09	0.22	0.22	0.22	0.22	0.22	0.22	0.19	0.22	0.08	0.53	0.19	0.02	0.02	0												
eri sarracenia	0.12	0.15	0.17	0.17	0.17	0.16	0.16	0.17	0.16	0.16	0.13	0.52	0.23	0.09	0.08	0.10	0											
lam byblis gigantea	0.10	0.15	0.16	0.16	0.16	0.16	0.17	0.16	0.13	0.17	0.14	0.52	0.21	0.12	0.13	0.12	0.09	0										
lam genlisea guianensis	0.59	0.60	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.60	0.55	0.60	0.60	0.60	0.60	0.59	0.57	0									
lam philcoxia minensis	0.15	0.12	0.23	0.23	0.23	0.23	0.24	0.23	0.21	0.23	0.11	0.53	0.21	0.08	0.08	0.08	0.15	0.12	0.59	0								
lam pinguicula pumila	0.12	0.15	0.17	0.17	0.17	0.17	0.17	0.17	0.16	0.17	0.14	0.51	0.24	0.13	0.13	0.13	0.07	0.07	0.58	0.13	0							
lam utricularia biflora	0.22	0.26	0.17	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.25	0.48	0.28	0.24	0.24	0.24	0.22	0.17	0.53	0.24	0.20	0						
oxa brocchinia micrantha	0.18	0.15	0.26	0.26	0.26	0.26	0.27	0.26	0.24	0.26	0.14	0.54	0.23	0.12	0.12	0.12	0.18	0.19	0.61	0.13	0.19	0.29	0					
poa catopsis subulata	0.21	0.18	0.28	0.29	0.29	0.29	0.29	0.29	0.26	0.29	0.17	0.55	0.26	0.15	0.15	0.15	0.21	0.21	0.61	0.15	0.22	0.31	0.06	0				
poa cephalotus	0.11	0.11	0.20	0.20	0.20	0.20	0.20	0.20	0.17	0.20	0.09	0.53	0.20	0.08	0.08	0.08	0.10	0.12	0.60	0.12	0.12	0.24	0.15	0.18	0			
x car	car	car	car	car	car	car	car	car	car	car	car	car	eri	eri	eri	eri	lam	lam	lam	lam	lam	oxa	poa	poa				

Bidirectional histogram

caryo aldrovanda vasiculosa	0																														
caryo dionaea muscipula	0.32	0																													
caryo drosera anglica	0.32	0.37	0																												
caryo drosera capillaris	0.33	0.39	0.05	0																											
caryo drosera filliformis	0.33	0.39	0.05	0.03	0																										
caryo drosera occidentalis	0.36	0.39	0.32	0.34	0.34	0																									
caryo drosera rosulata	0.40	0.40	0.35	0.36	0.37	0.27	0																								
caryo drosera rotundifolia	0.32	0.38	0.02	0.07	0.07	0.32	0.35	0																							
caryo drosera spathulata	0.31	0.37	0.07	0.08	0.10	0.33	0.35	0.08	0																						
caryo drosera tokaiensis	0.32	0.38	0.05	0.06	0.08	0.32	0.34	0.06	0.01	0																					
caryo drosophyllum lusitancium	0.36	0.38	0.39	0.40	0.40	0.40	0.43	0.39	0.38	0.38	0																				
caryo nepenthes vieillardii	0.82	0.83	0.84	0.84	0.84	0.85	0.85	0.84	0.83	0.84	0.80	0																			
caryo triphyophyllum peltatum	0.41	0.43	0.46	0.47	0.46	0.46	0.47	0.46	0.45	0.45	0.34	0.82	0																		
eri darlingtonia californica	0.48	0.47	0.49	0.48	0.49	0.48	0.50	0.49	0.47	0.47	0.40	0.80	0.42	0																	
eri heliamphora ionasii	0.47	0.45	0.48	0.48	0.48	0.49	0.49	0.49	0.47	0.47	0.41	0.80	0.42	0.12	0																
eri roridula gorgonias	0.47	0.44	0.48	0.48	0.49	0.48	0.49	0.48	0.46	0.46	0.41	0.78	0.43	0.12	0.16	0															
eri sarracenia	0.49	0.47	0.49	0.48	0.49	0.48	0.47	0.49	0.48	0.47	0.45	0.83	0.46	0.20	0.12	0.21	0														
lam byblis gigantea	0.45	0.48	0.47	0.46	0.46	0.47	0.50	0.48	0.45	0.46	0.43	0.83	0.46	0.37	0.37	0.36	0.40	0													
lam genlisea guianensis	0.68	0.69	0.66	0.66	0.66	0.66	0.68	0.66	0.66	0.66	0.66	0.84	0.68	0.64	0.65	0.64	0.66	0.56	0												
lam philcoxia minensis	0.50	0.50	0.52	0.52	0.51	0.51	0.54	0.52	0.50	0.52	0.48	0.81	0.51	0.42	0.42	0.40	0.44	0.29	0.62	0											
lam pinguicula pumila	0.52	0.50	0.51	0.50	0.50	0.51	0.51	0.52	0.50	0.50	0.46	0.85	0.49	0.39	0.37	0.38	0.39	0.29	0.59	0.36	0										
lam utricularia biflora	0.51	0.55	0.52	0.50	0.50	0.51	0.55	0.53	0.51	0.52	0.49	0.81	0.50	0.43	0.44	0.45	0.47	0.33	0.57	0.43	0.38	0									
oxa brocchinia micrantha	0.58	0.54	0.57	0.57	0.57	0.58	0.58	0.57	0.57	0.57	0.55	0.79	0.56	0.47	0.47	0.48	0.49	0.52	0.68	0.53	0.52	0.55	0								
poa catopsis subulata	0.58	0.55	0.57	0.57	0.57	0.58	0.58	0.58	0.57	0.57	0.55	0.79	0.56	0.48	0.48	0.49	0.49	0.50	0.68	0.52	0.51	0.54	0.13	0							
poa cephalotus	0.48	0.46	0.50	0.49	0.50	0.50	0.52	0.50	0.48	0.48	0.43	0.82	0.45	0.35	0.35	0.34	0.37	0.43	0.68	0.46	0.42	0.48	0.50	0.50	0						
x car	car	car	car	car	car	car	car	car	car	car	car	car	eri	eri	eri	eri	lam	lam	lam	lam	lam	oxa	poa	poa							

Data analysis

- It turns out that manual analysis of the tree similarity is almost impossible.
- The best results are given by the compression algorithm and then the surface dotplot area algorithm. The dotplot algorithm actually seems to give a bit more consistent results.
- Dotplot area algorithm and histogram give quite similar results. The results of histogram are a bit worse.
- Levenstein distance gives quite garbled trees.
- Graphical dotplots show that similarity of samples is exceptional (approximately 70% of DNA has 100% correspondence with window of length 20).
- As the consequence, uni and bi directional versions of the histogram give exactly the same results.
- Nepenthes family results turn out to be crucial for result analysis. Unfortunately, the only present sample does not match its expected position in any of the trees.

Data analysis

- The phylogenetic trees mostly correspond to those the original paper proposes.
- Also, the sample corresponds to the standard taxonomy.
- The data sample is too small to draw other conclusion with respect to evolution of trap types.
- Our sample contains some genes which were not available to the original study. On the other hand, the original study contains much more sophisticated system of representants of the Caryphyllales order. This results in an effective imposibility of drawing exact results.
- GenCompress is unable digest Drosera Spathulata for some reason.

- The original study: http://www.botany.wisc.edu/courses/botany_400/papers/Carnivorou
- http://www.carnivorousplants.org/cp/WhatAreCPs.php
- http://journals.plos.org/plosone/article?id=10.1371/journal.pone.00
- https://www.czplants.com
- http://www.masozravky.com/
- http://botany.org/Carnivorous_Plants/