Identifying reef habitat states using machine learning applied to the global *Reef Life Survey* dataset

Clément Violet[1], Aurélien Boyé[1], Rick Stuart-Smith[2], Graham Edgar[2], Martin Marzloff[1]

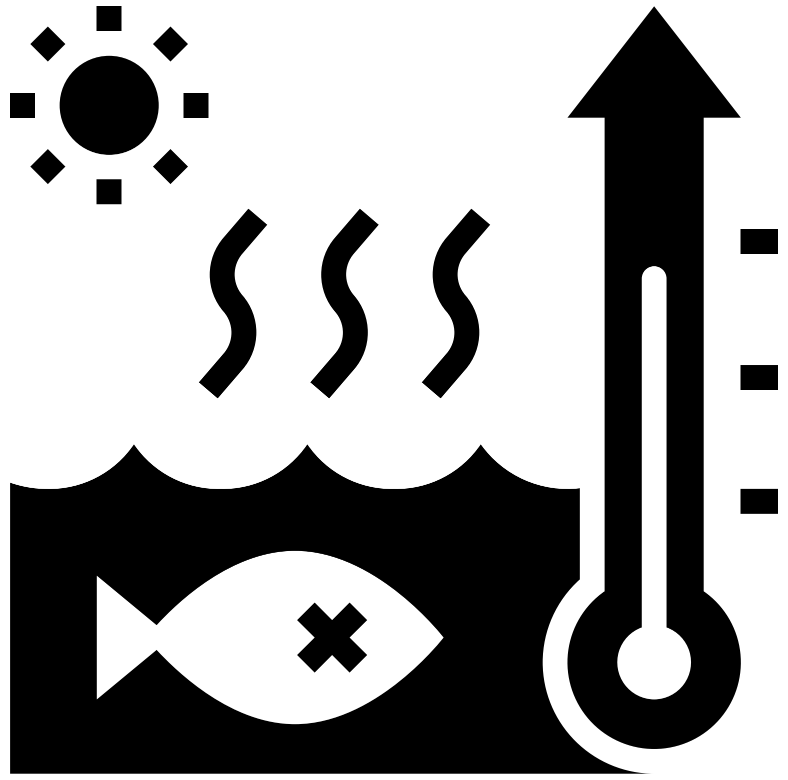
1. IFREMER, DYNECO LEBCO, Plouzané 29280, France

2. Institute for Marine and Antarctic Studies, University of Tasmania, Hobart 7001, Australia

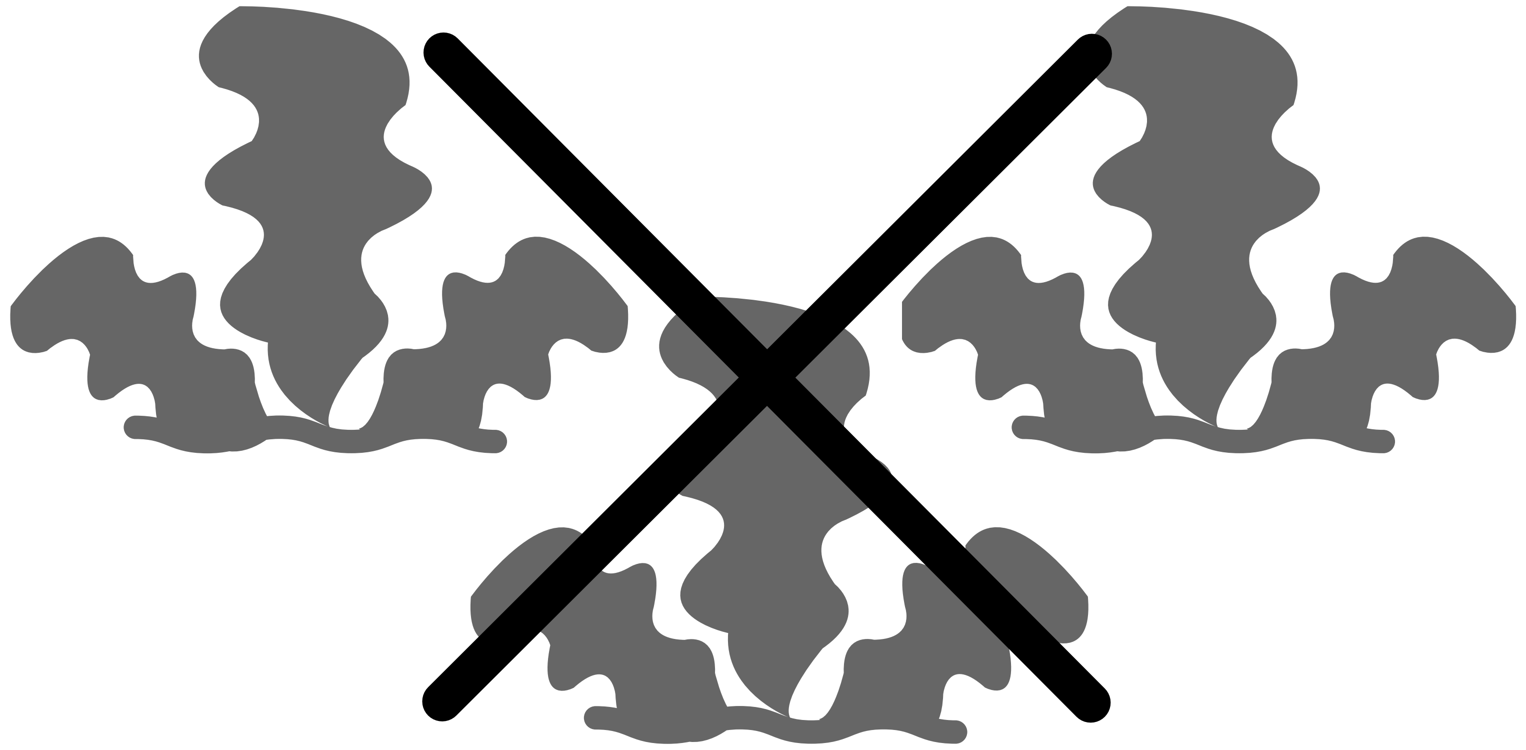
©Picture by John Turnbull

Introduction

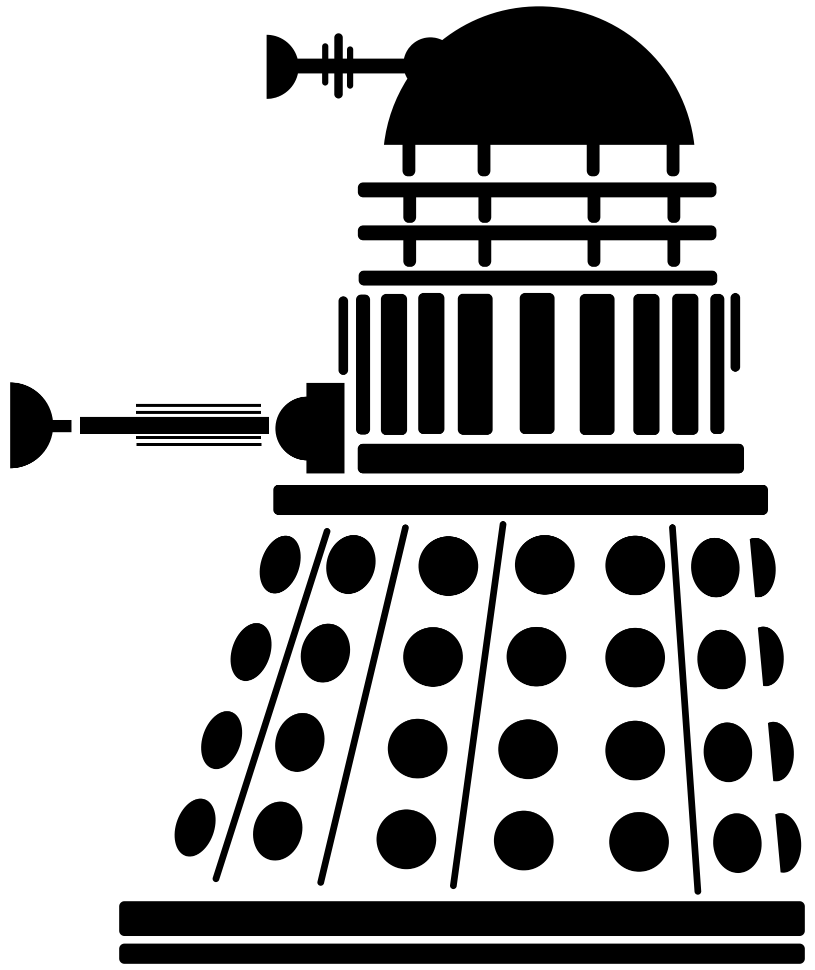
Coastal ecosystems are affected by multiple stressors



Climate change

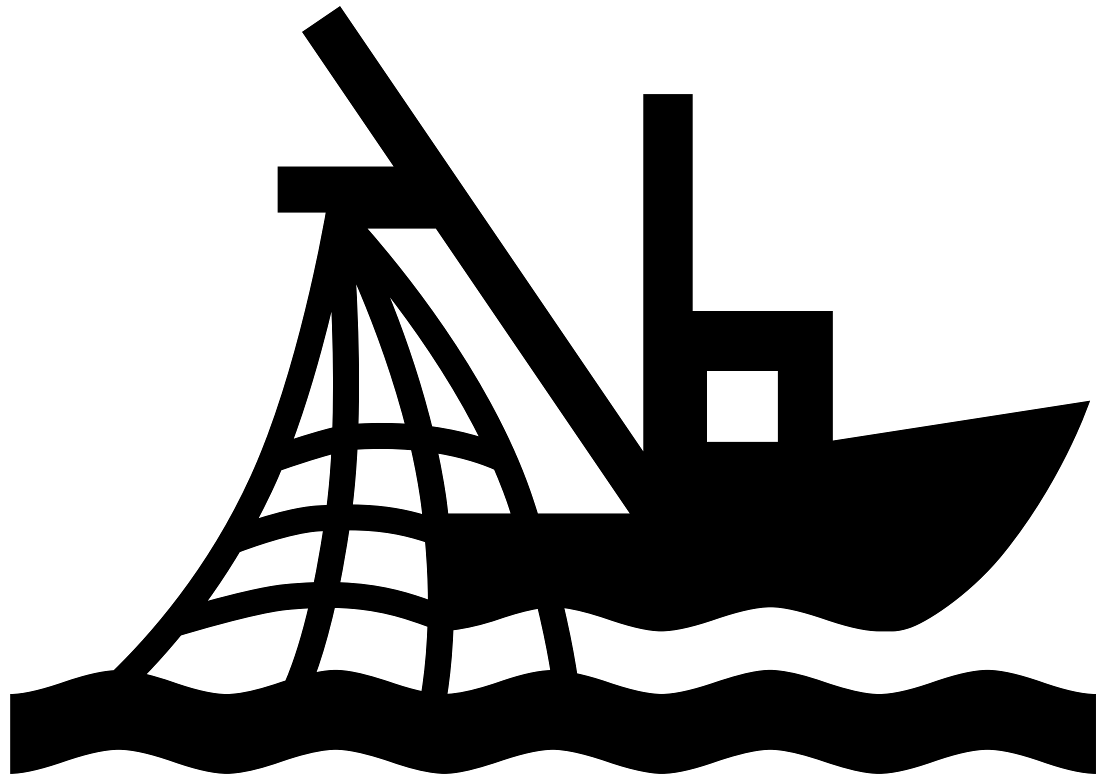


Habitat loss



Invasive species

Pollution



Overfishing

A typical exemple of alternative ecological states

Marine forests

©Picture by John Turnbull

- Complex & heterogenous tri-dimensional habitats
- High species richness
- High persistence

Sea-urchins barrens

©Picture by John Turnbull

- Homegenous and low complexity habitats
- Low species diversity
- High persistence

Bernal-Ibáñez *et al.* 2021

Challenges of identifying alternative ecological states

1. Complexity of ecosystems:
    - Many components and processes maintain an ecosystem in a given state

©Picture by John Turnbull

Challenges of identifying alternative ecological states

1. Complexity of ecosystems:
2. Variability of ecosystems:
   - Ecosystems are dynamics and can exhibit natural fluctuations
   - Sparse observations

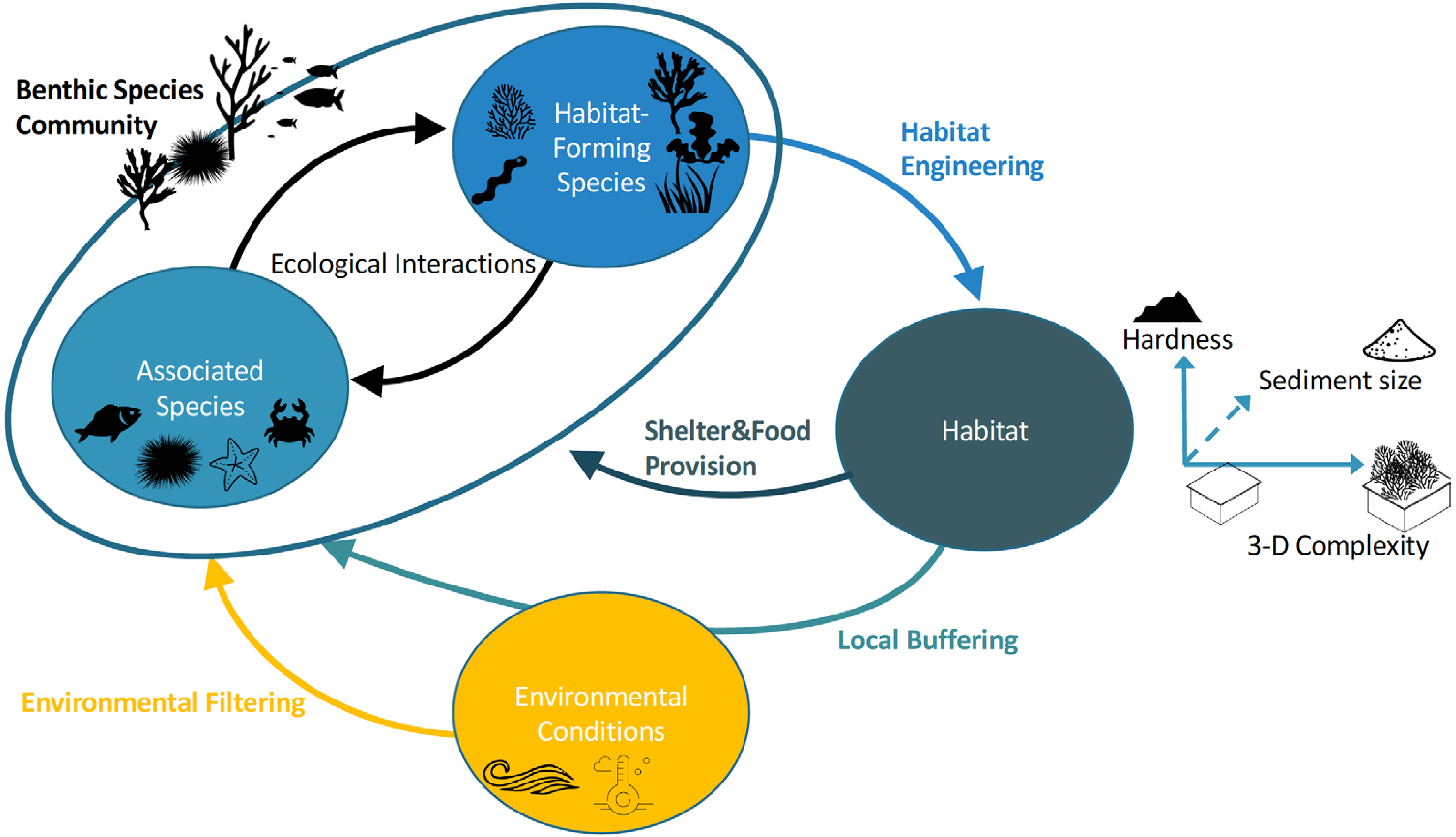
©Picture by John Turnbull

Challenges of identifying alternative ecological states

1. Complexity of ecosystems
2. Variability of ecosystems
3. Changes between alternative states can be fast and non-linear
    - Difficulties to identify threshold and/or observe transitions

©Picture by John Turnbull

A need for a systematic assessment of dominant benthic habitat states



Dominant habitat types can serve as proxys for ecological states on reef systems


©Picture by John Turnbull


©Picture by John Turnbull

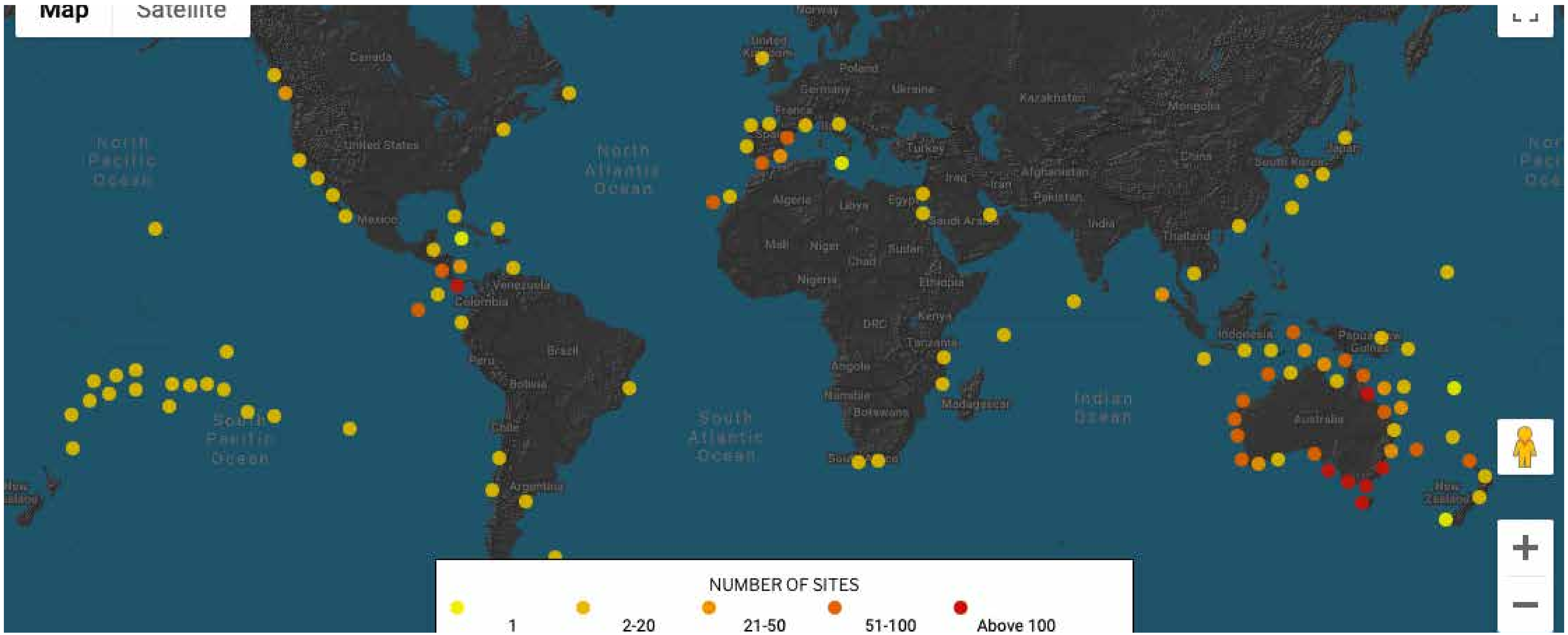
A need for a systematic assessment of dominant benthic habitat states

1. **Data-driven identification of dominant benthic habitat states at a global scale**
2. Characterise spatial distribution of dominant benthic habitat states and quantify influence environmental conditions and anthropogenic impacts (across global gradients)

Material & Methods

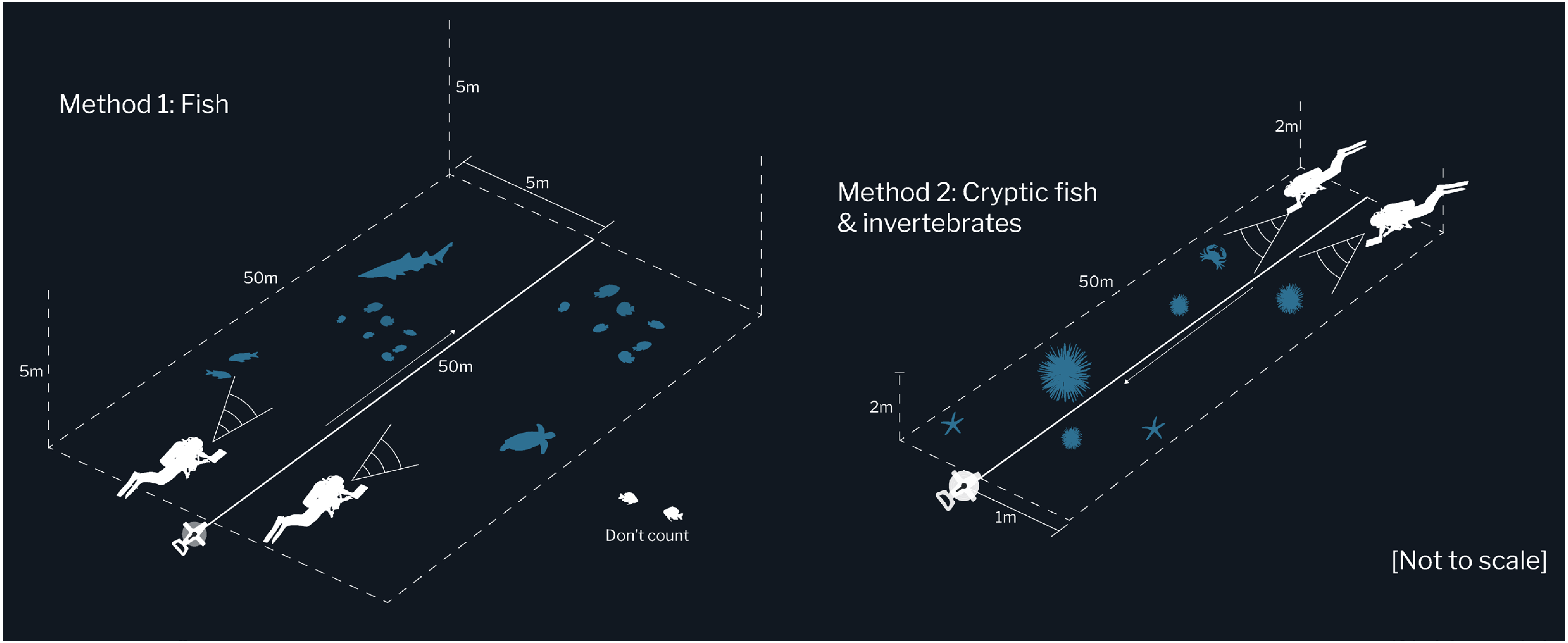
Reef Life Survey Program

6642 sites around the world



©Picture by John Turnbull

Reef Life Survey Program



Reef Life Survey Program

Photoquadrats every 2.5m

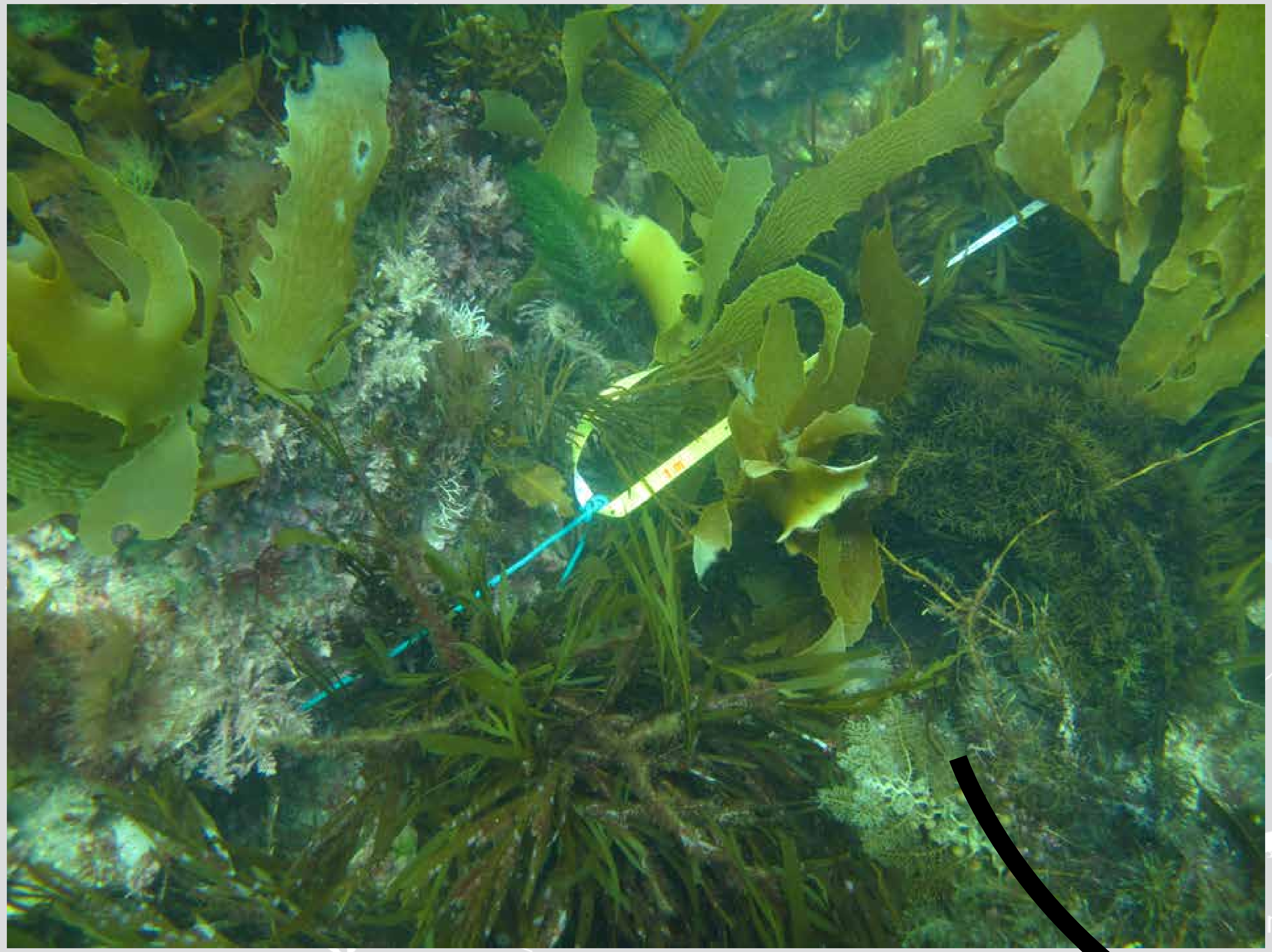
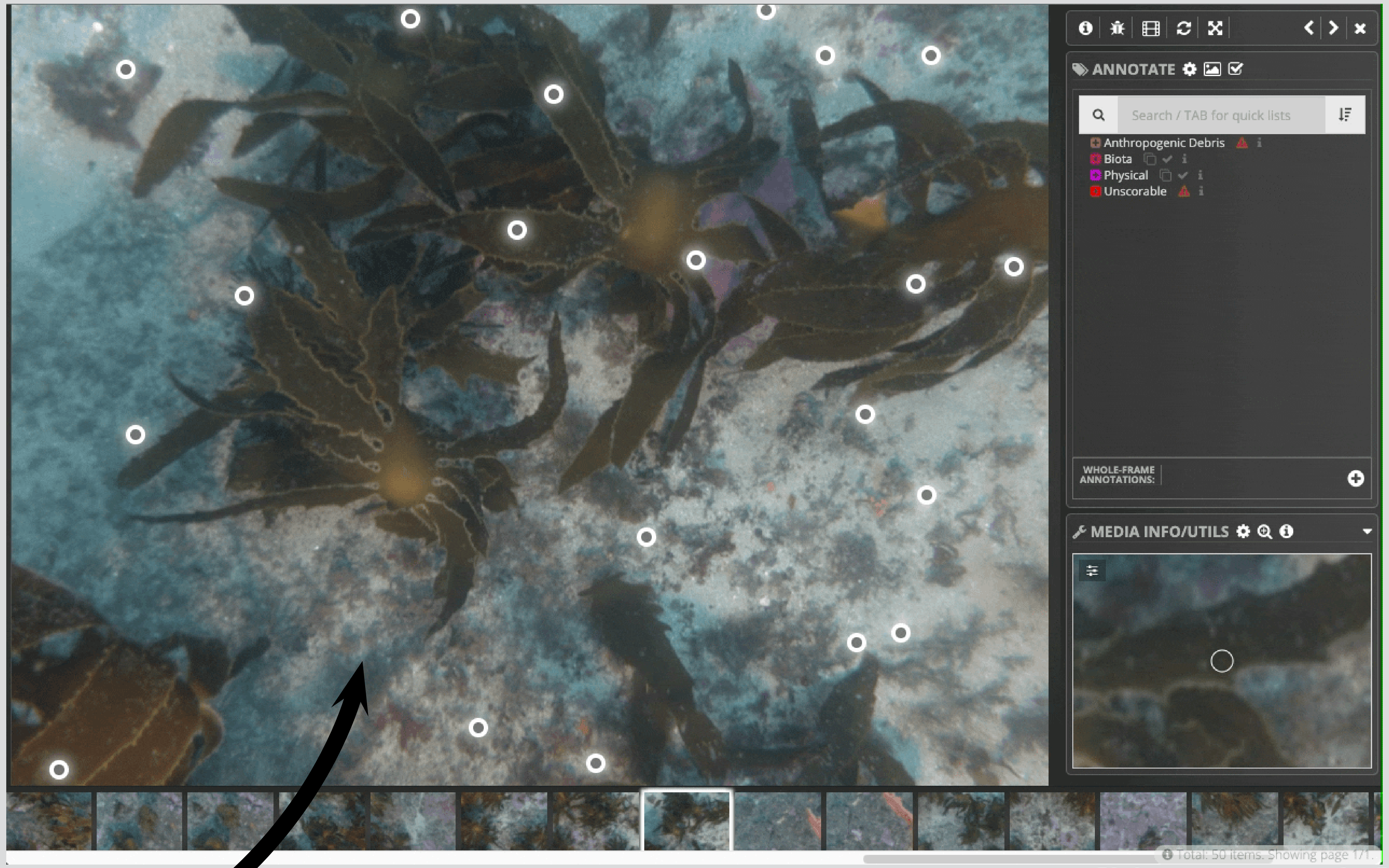


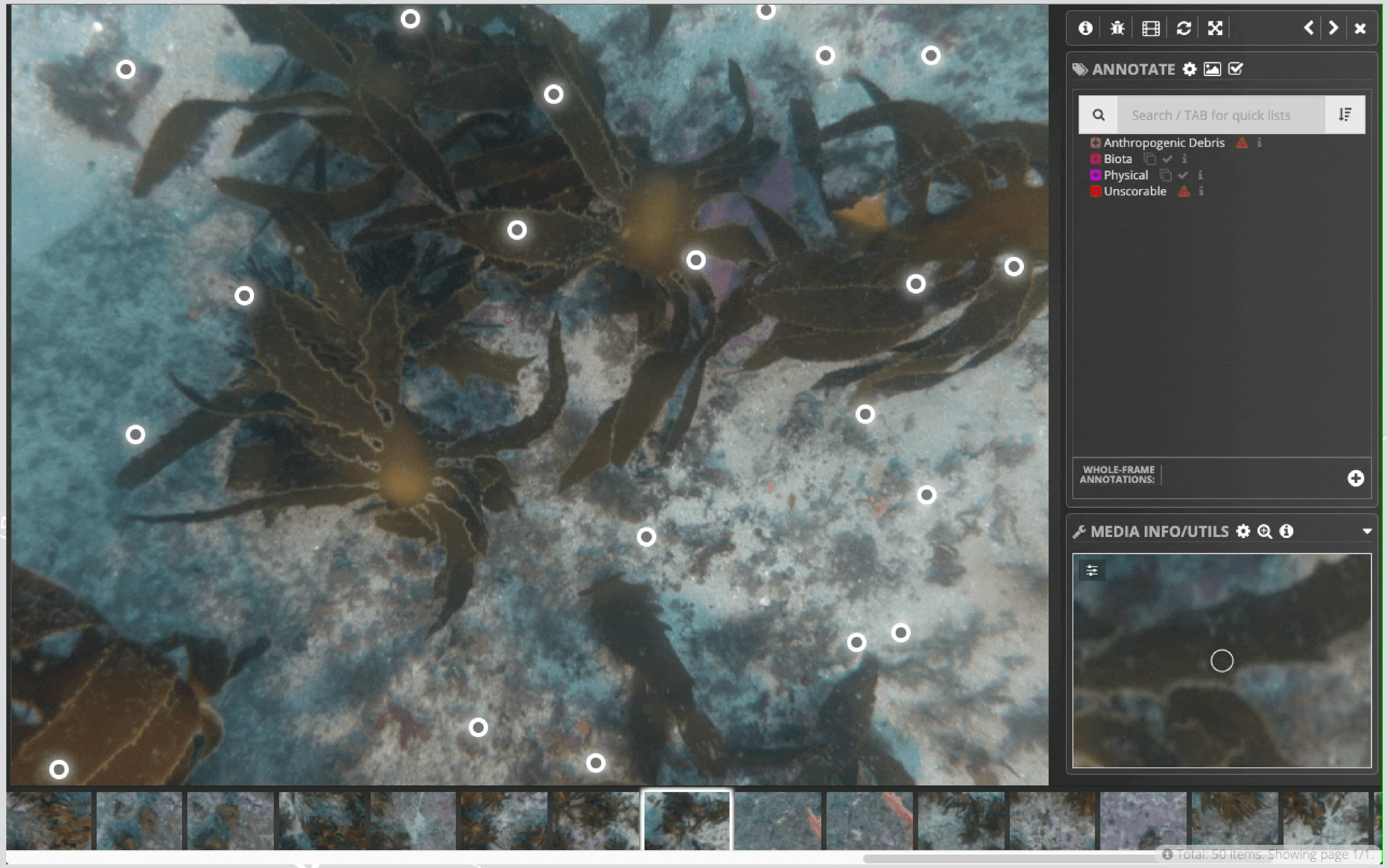
Image annotation on Squidle+



[Not to scale]

Reef Life Survey Program

Image annotation on Squidle+



% Cover 15 habitat types

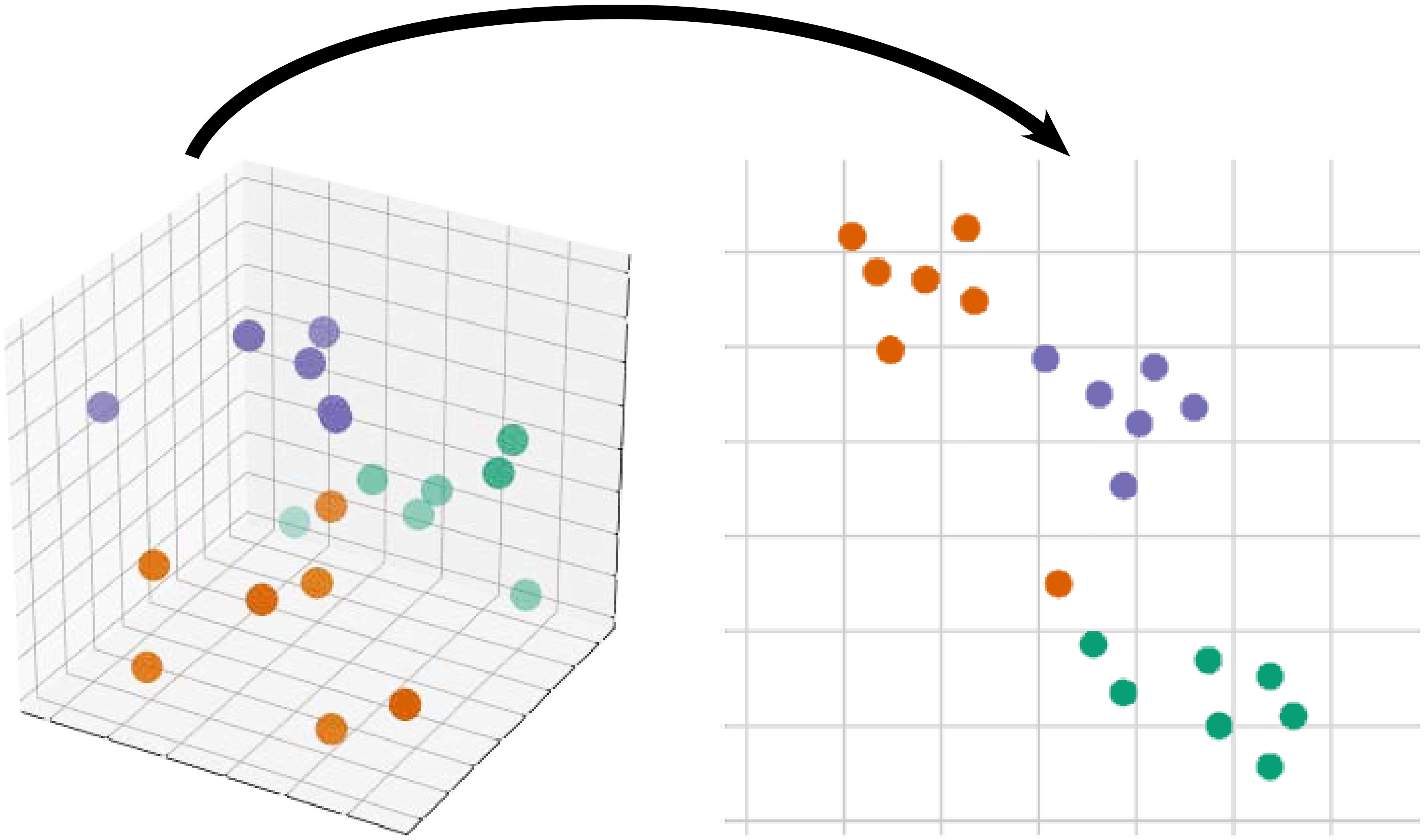
- Seagrasses
- Kelps
- Fucoids
- Turf alge
- Green algae
- Brown algae
- Red algae
- Filamentous epiphytes
- Encrusting algae
- Calcified algae
- Corals
- Hydrocorals
- Sessile invertebrates
- Non-living
- Sand

[Not to scale]

Clustering pipeline

Dimension reduction

UMAP (*McInnes et al.*, 2020)

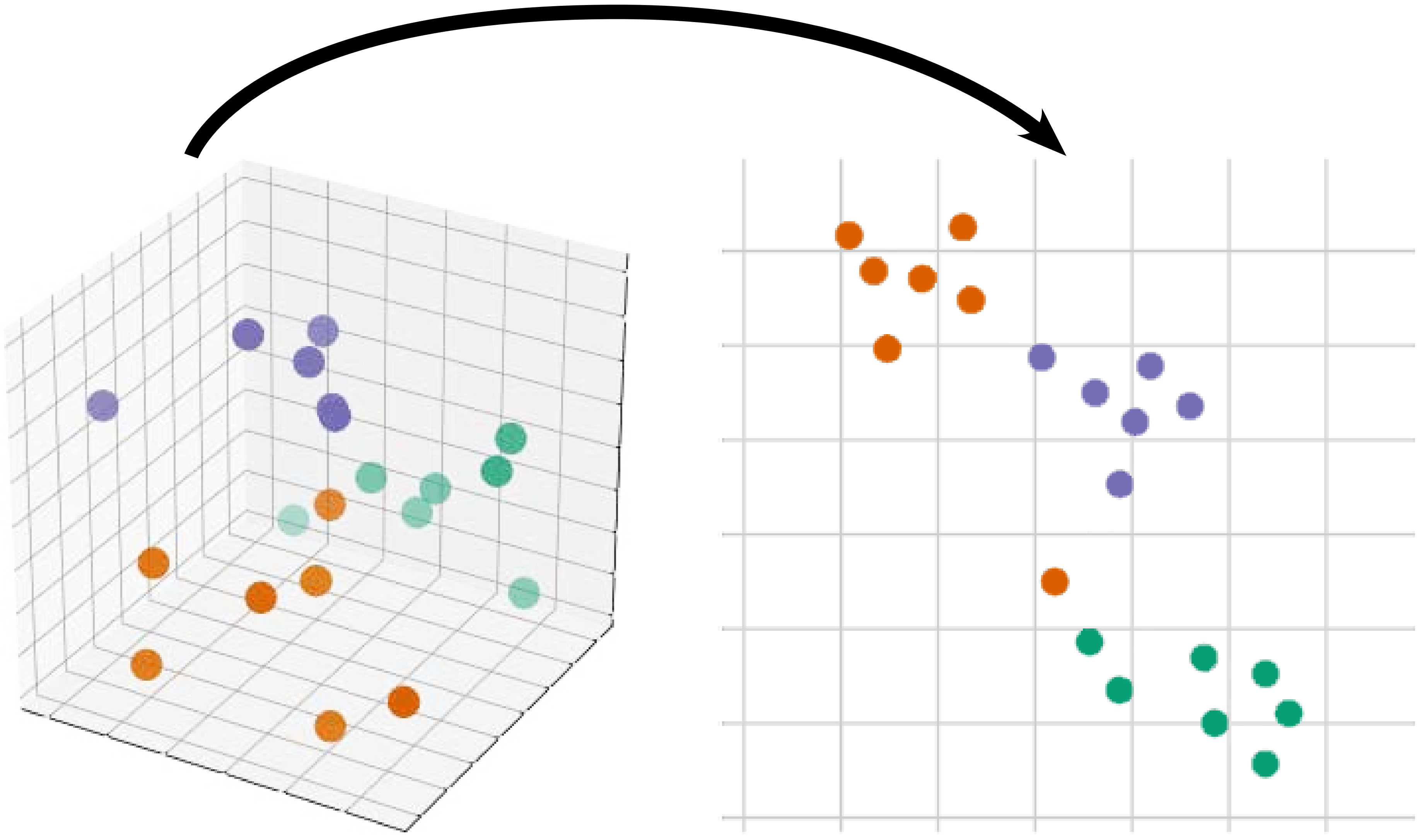


©Picture by John Turnbull

Clustering pipeline

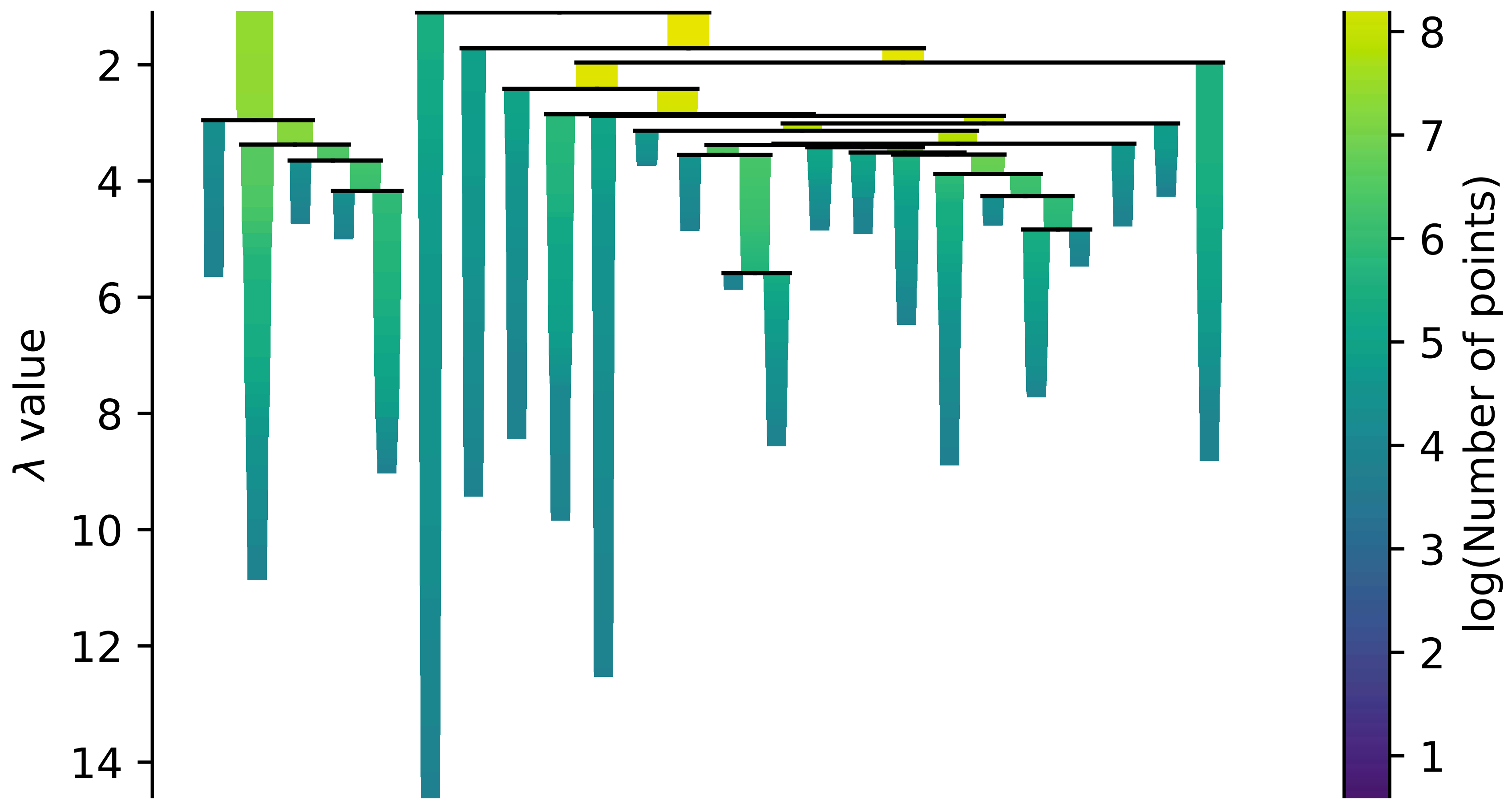
Dimension reduction

UMAP (*McInnes et al.*, 2020)



Clustering

HDBSCAN (*McInnes and Healy*, 2017)

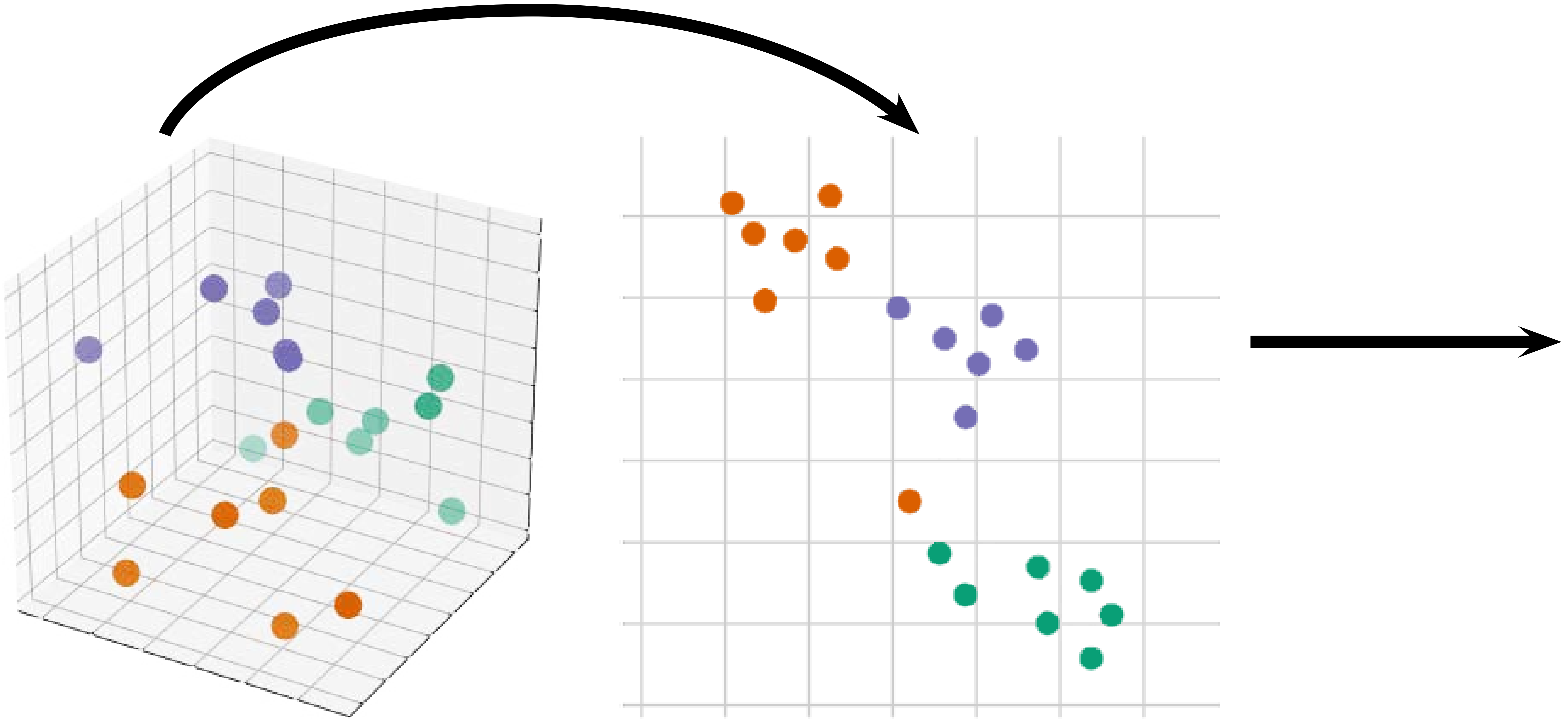


©Picture by John Turnbull

Clustering pipeline

Dimension reduction

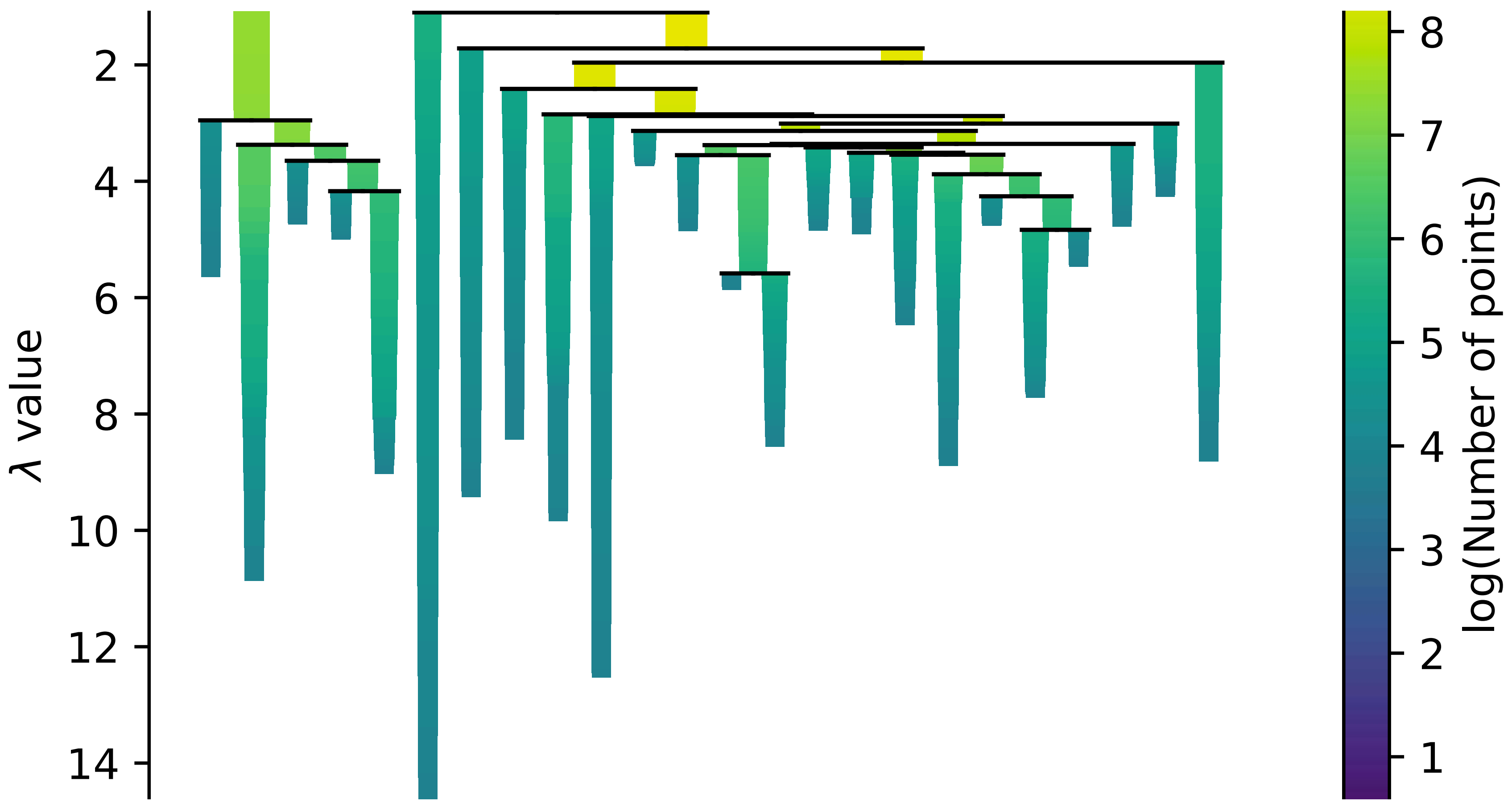
UMAP (*McInnes et al.*, 2020)



- Non-linear relationship
- Robust against noise and outliers
- Cluster of variable shapes & density

Clustering

HDBSCAN (*McInnes and Healy*, 2017)

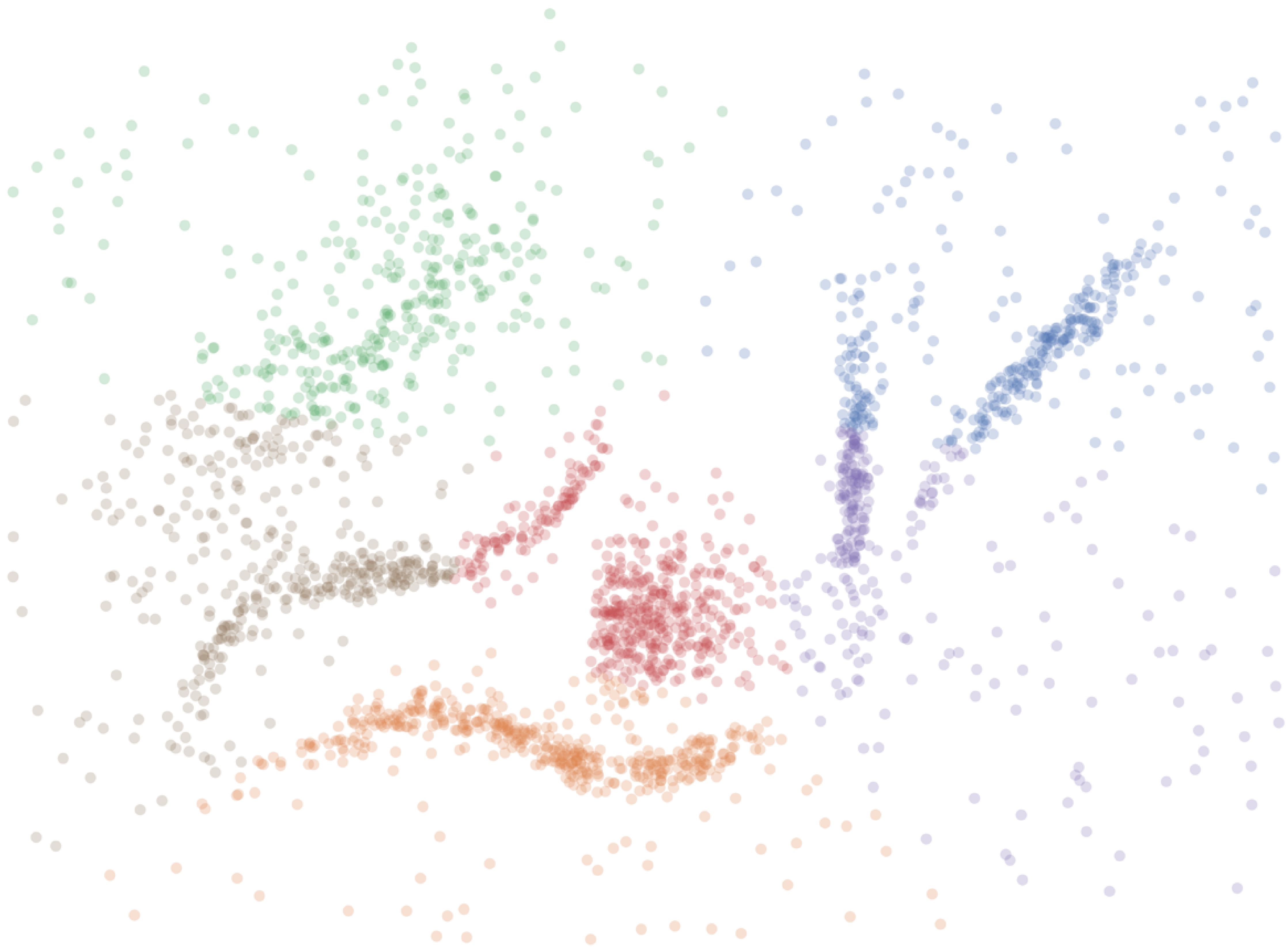


- Identification of representative members of each cluster

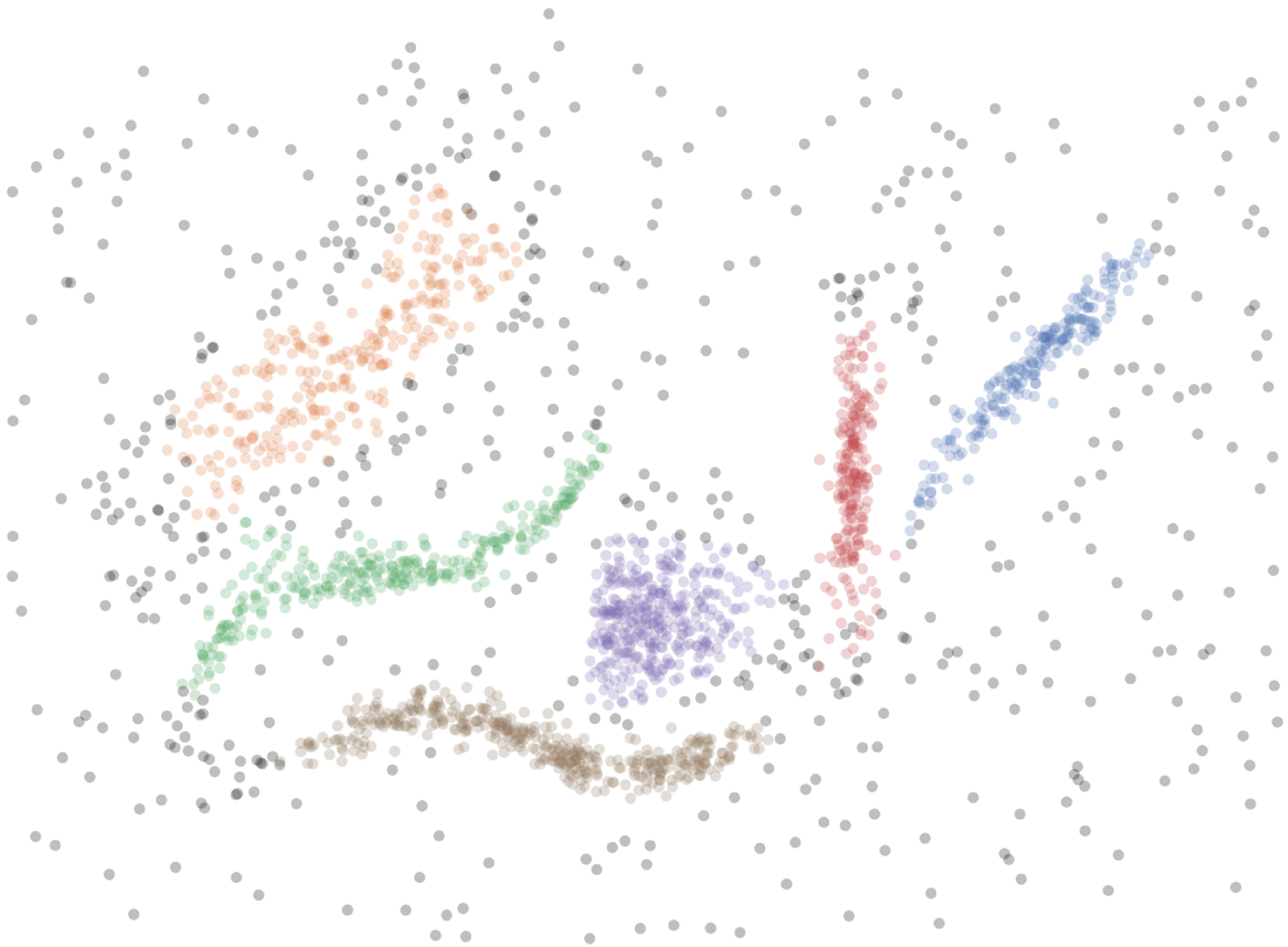
©Picture by John Turnbull

Comparison of clustering techniques on an example dataset

K-means



HDBSCAN

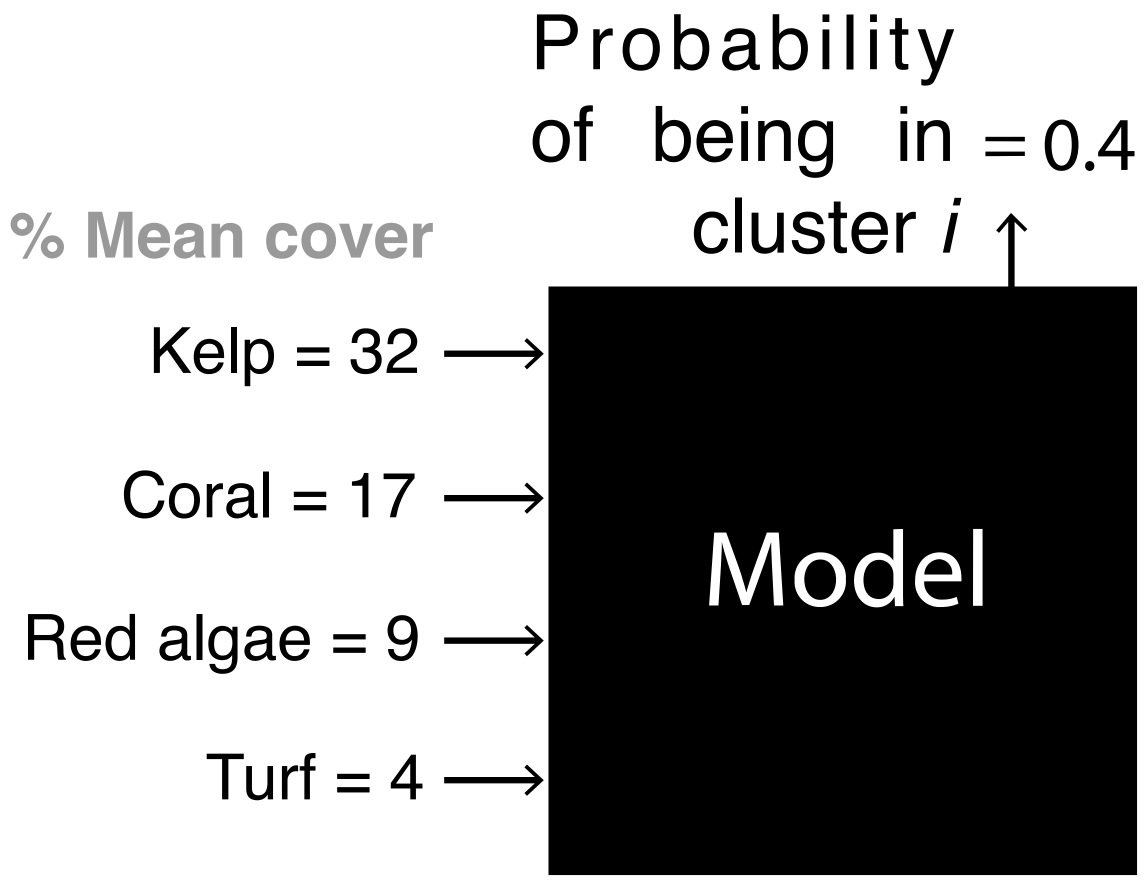



©Picture by John Turnbull

Interpretation of clustering results

- Complex pipeline can act as ‘black-boxes’
  - SHAP framework to understand probabilities of belonging to clusters *(Lunderberg et al.*, 2017)

SHAP

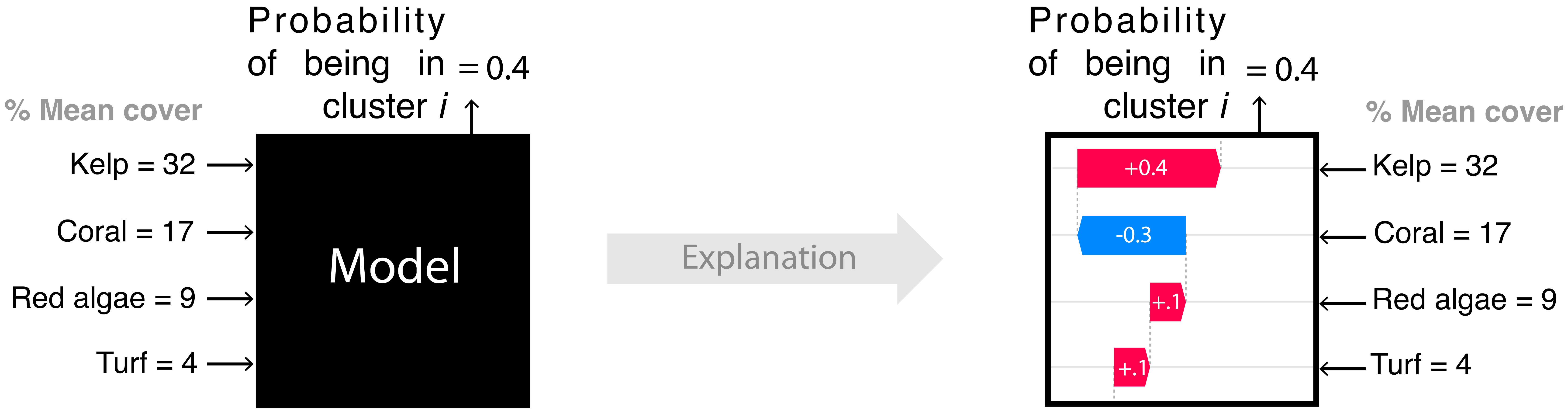



©Picture by John Turnbull

Interpretation of clustering results

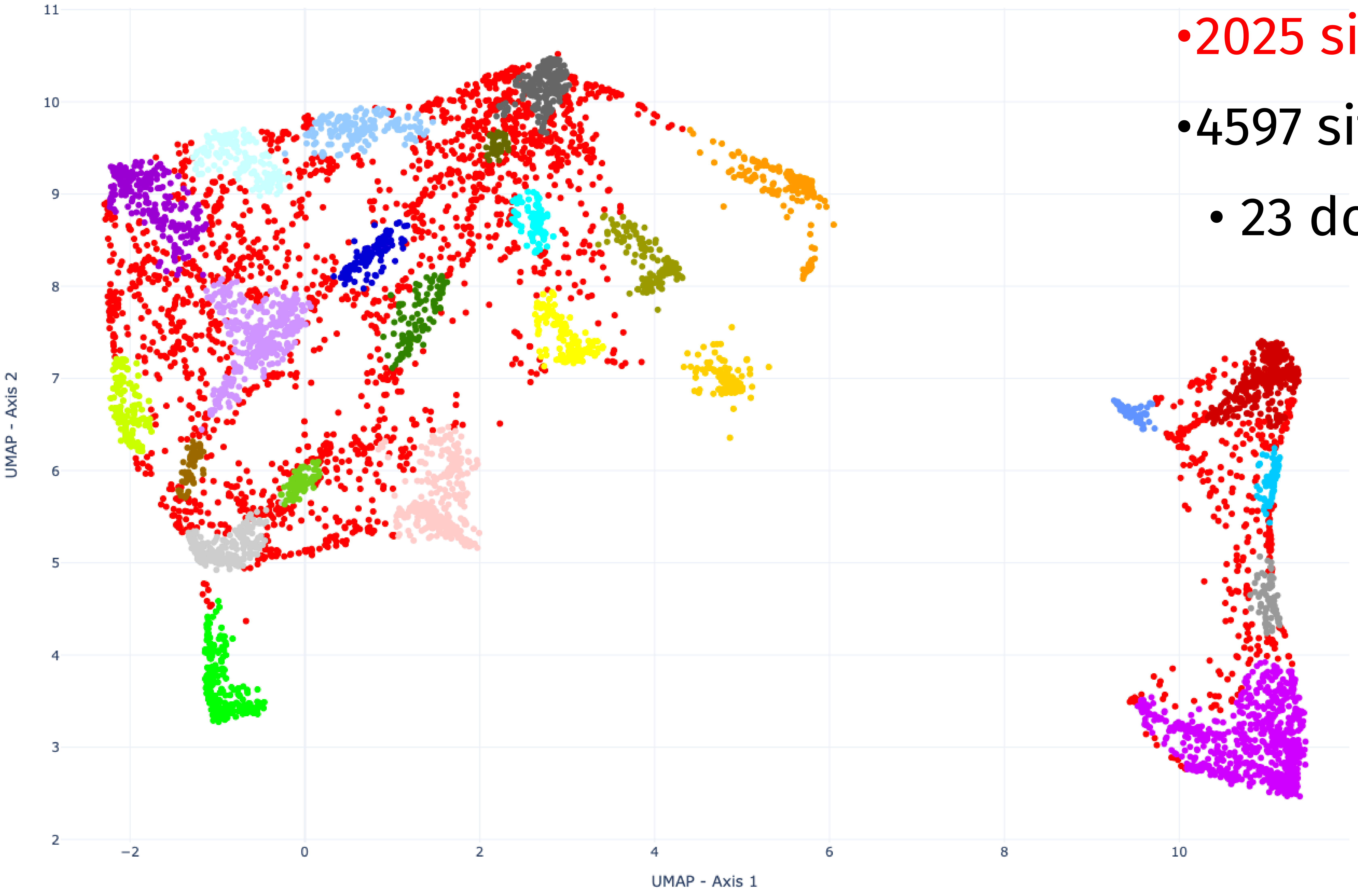
- Complex pipeline can act as ‘black-boxes’
  - SHAP framework to understand probabilities of belonging to clusters *(Lunderberg et al.*, 2017)

SHAP



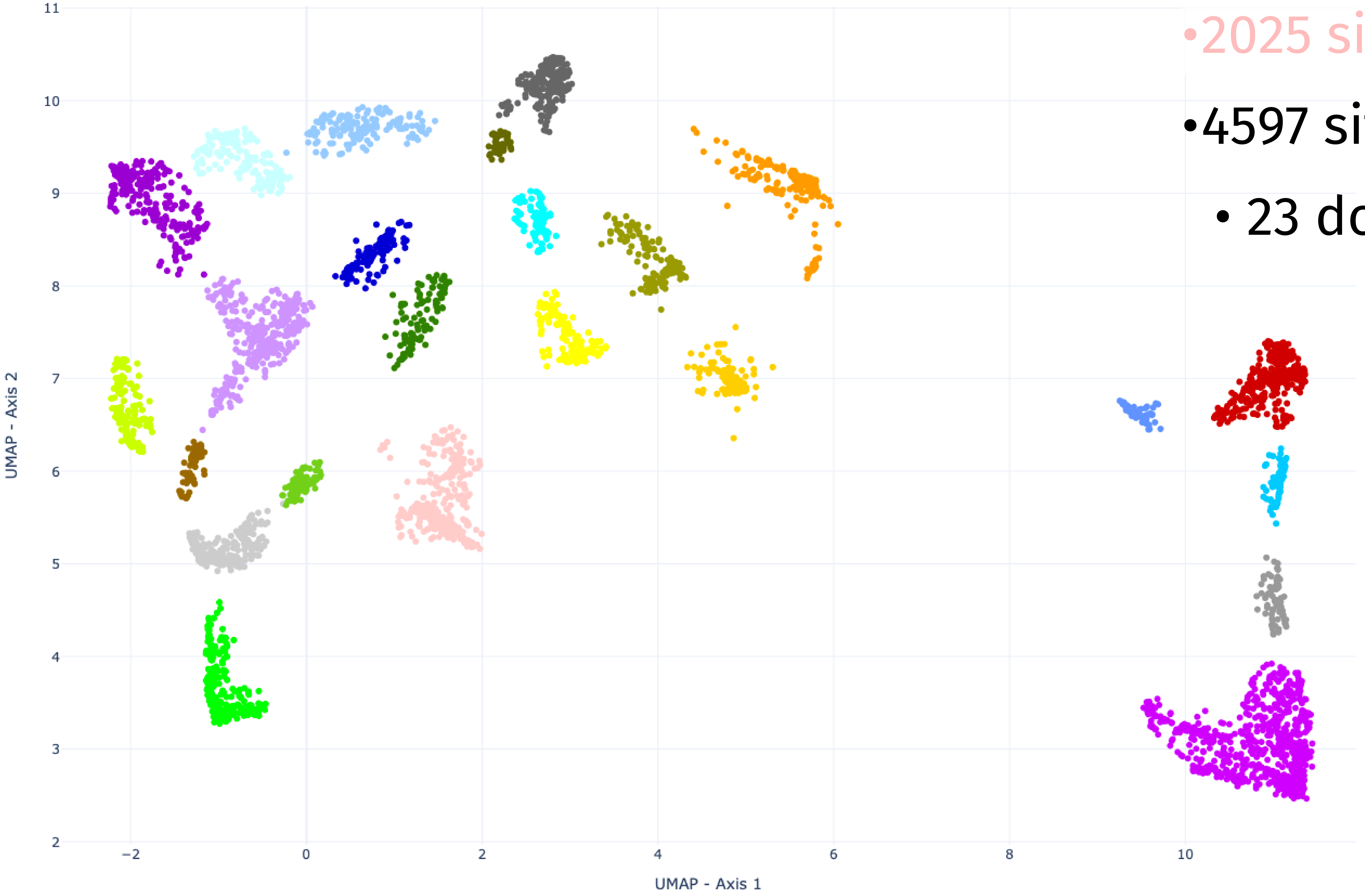
Results

Clustering *RLS* transects



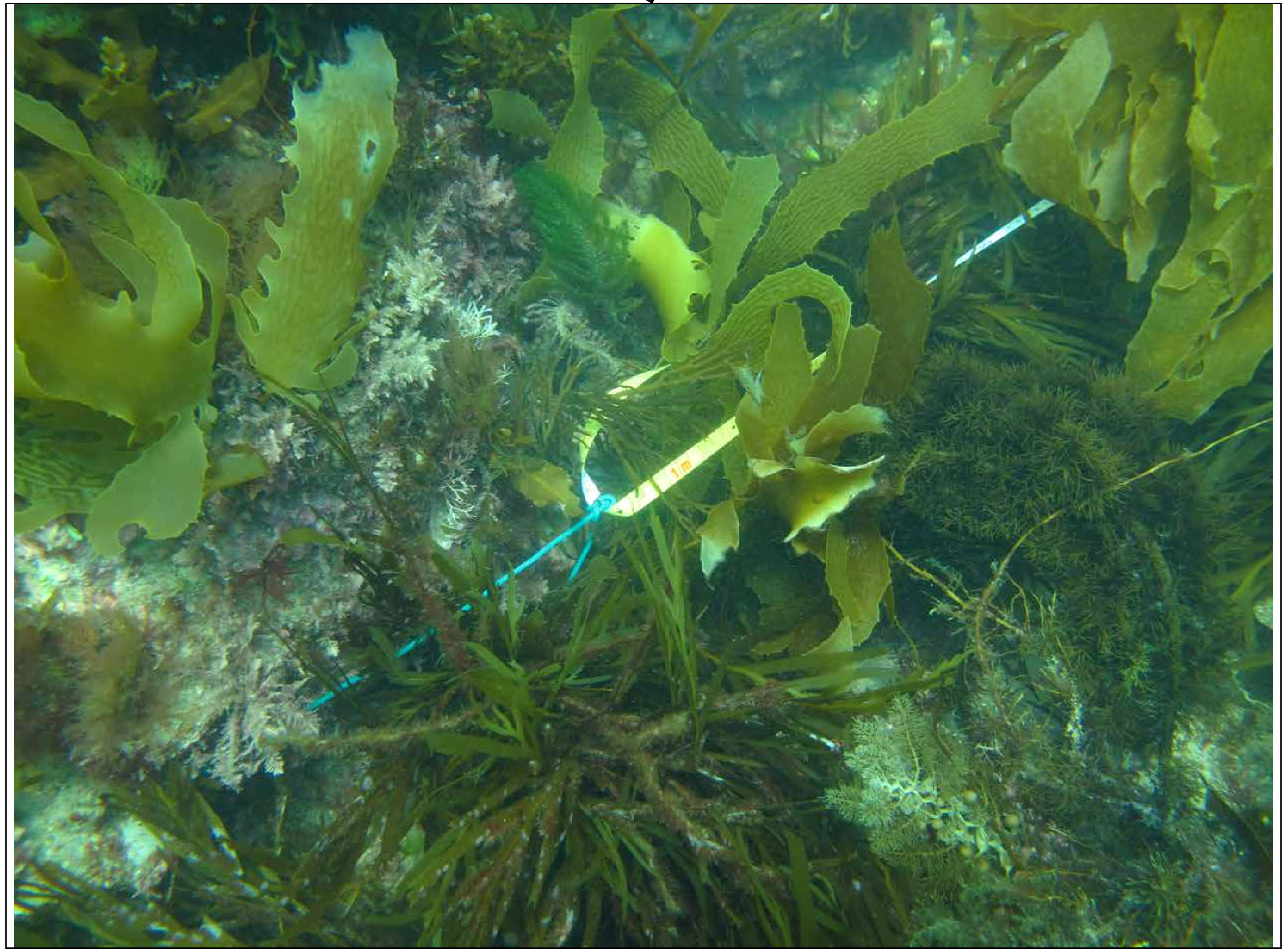
- 2025 sites classified as noise
- 4597 sites classified
  - 23 dominant habitat types

Clustering *RLS* transects

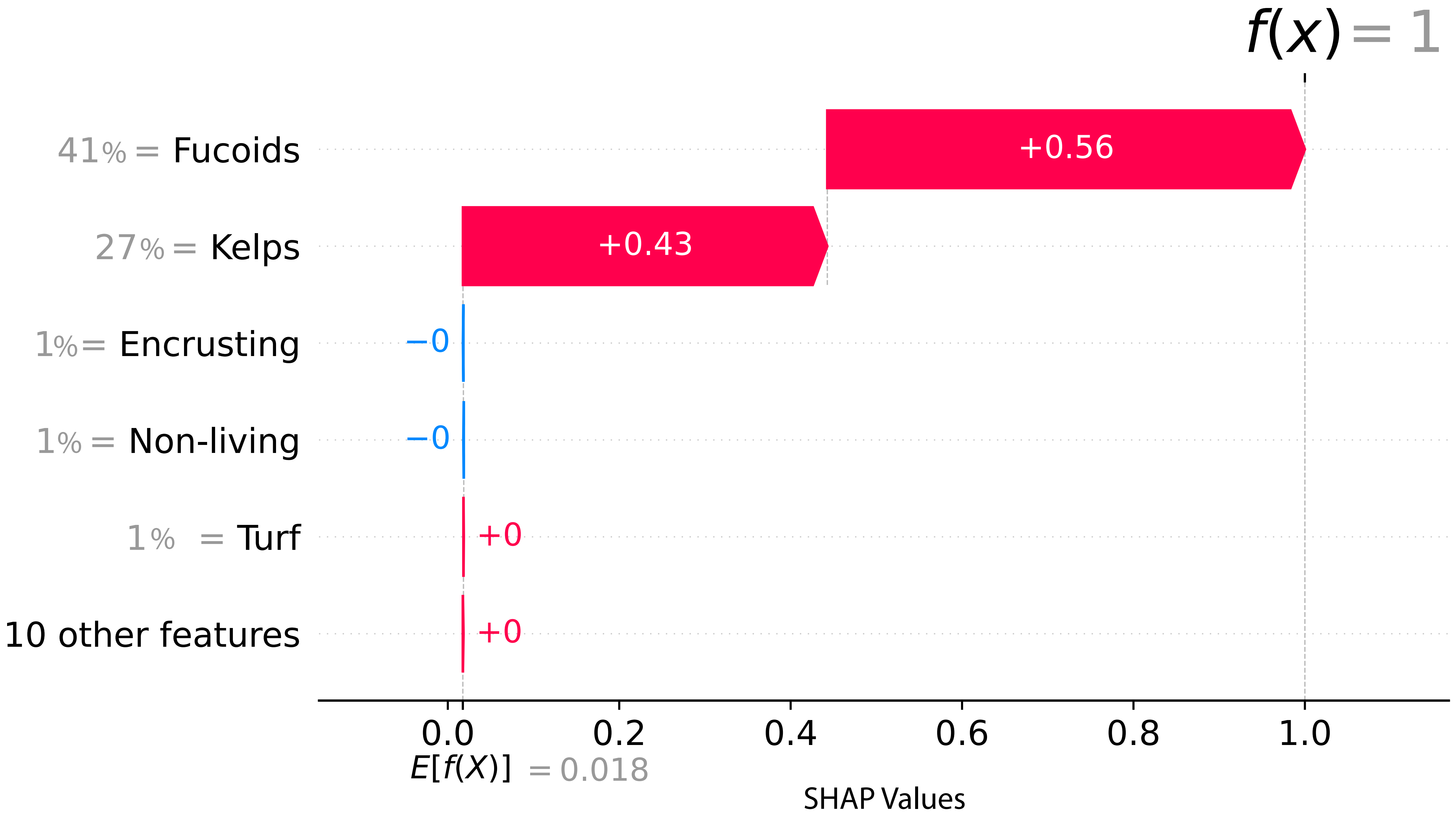


- 2025 sites classified as noise
- 4597 sites classified
  - 23 dominant types

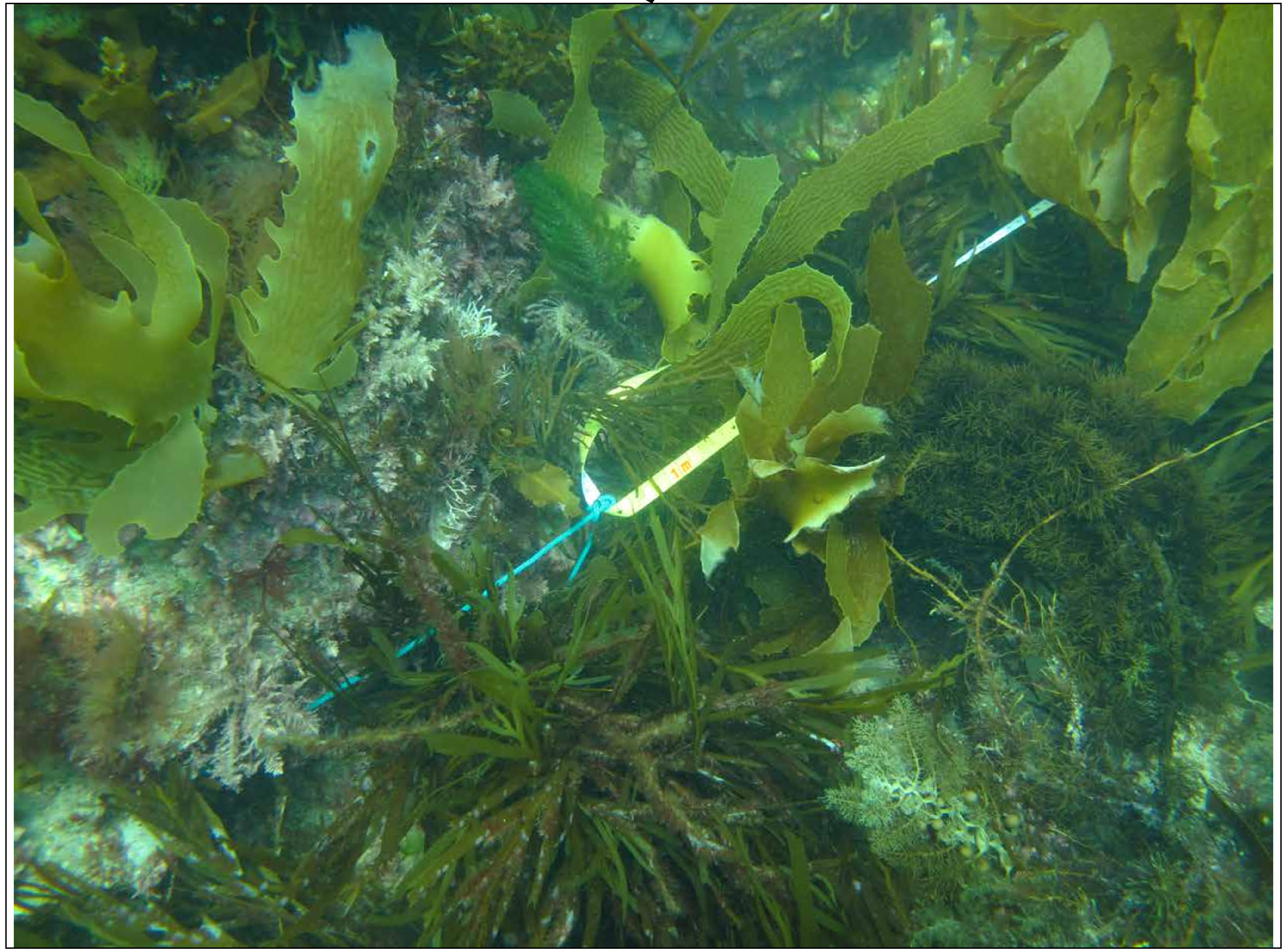
Explaining one member of a cluster



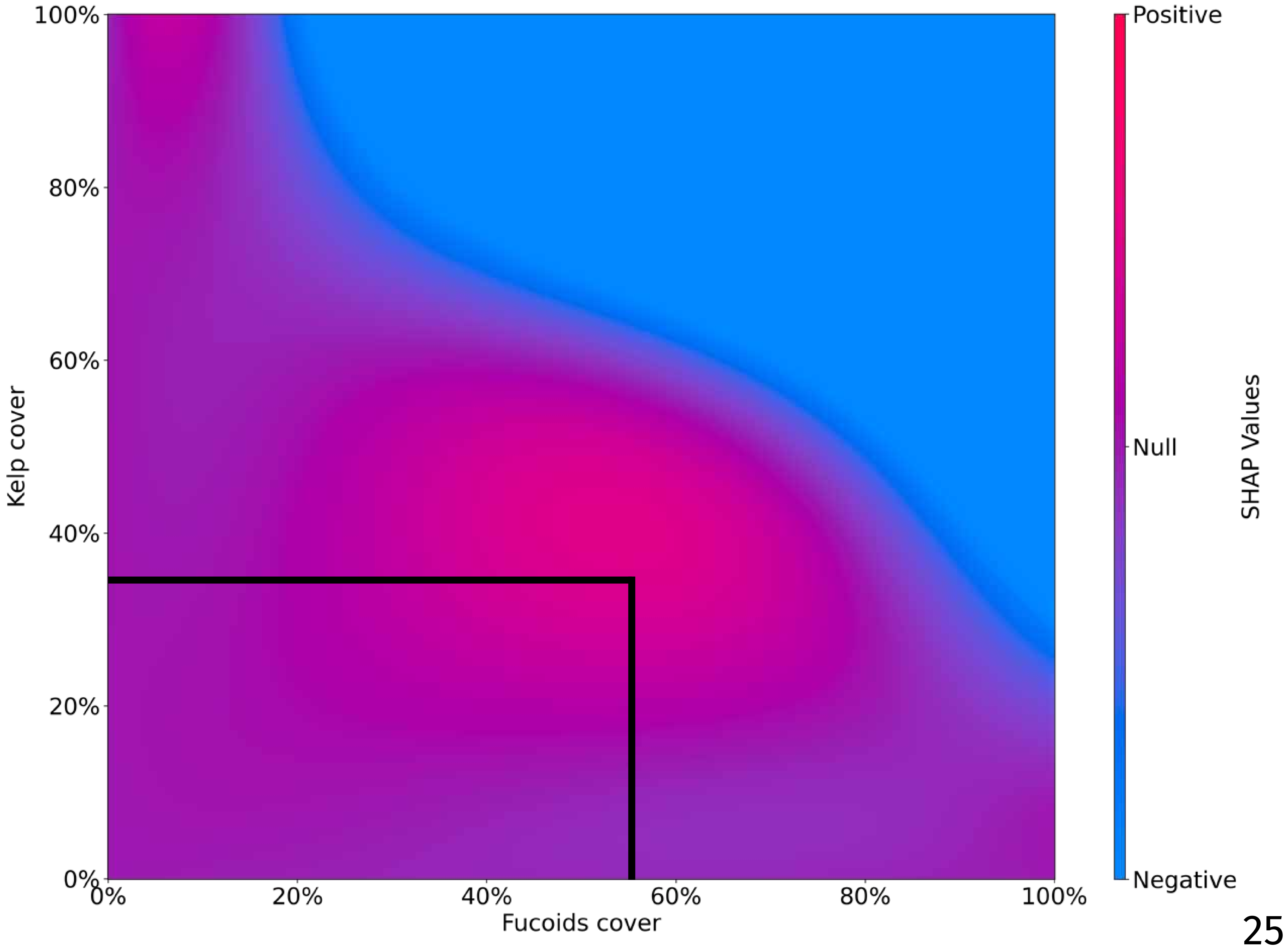
Relative contribution of each variable to this cluster



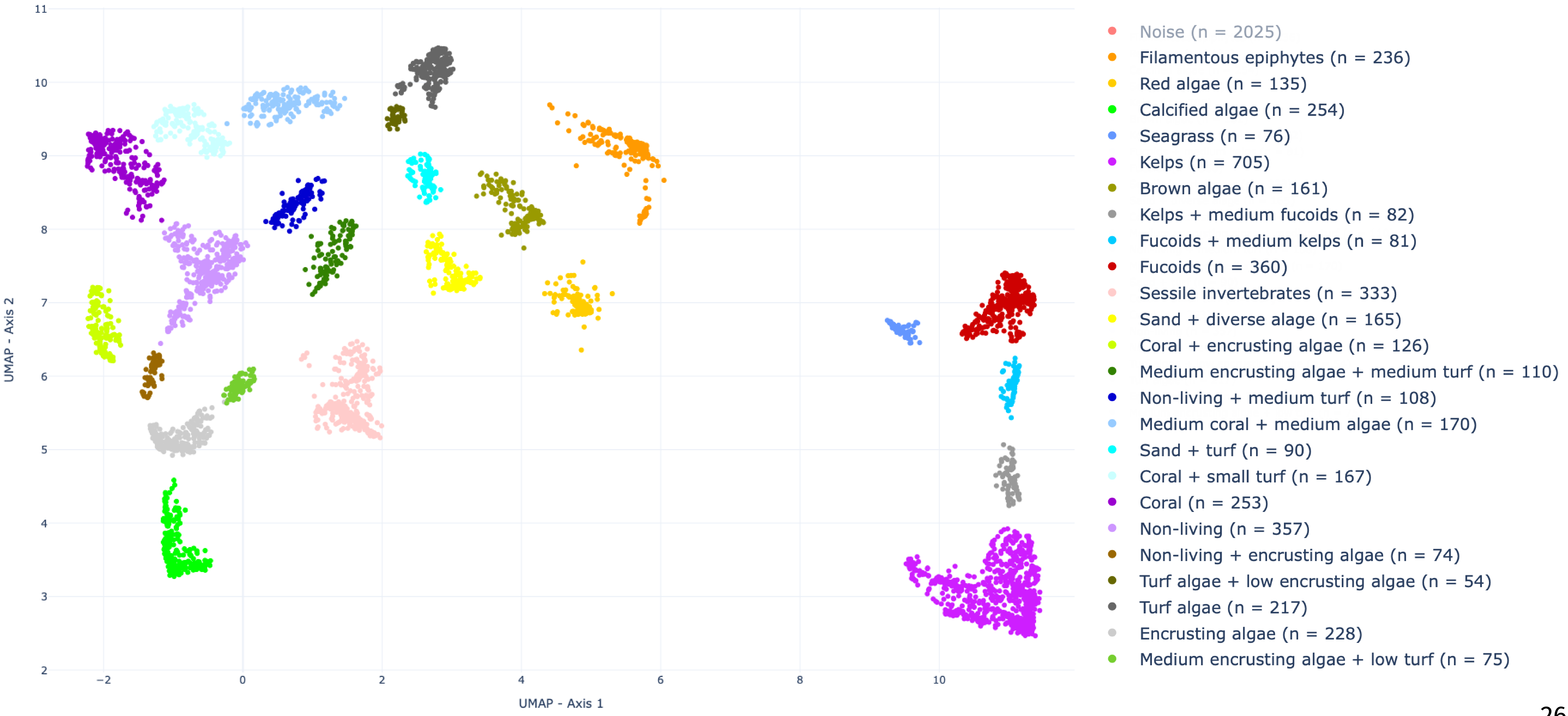
Explaining one member of a cluster



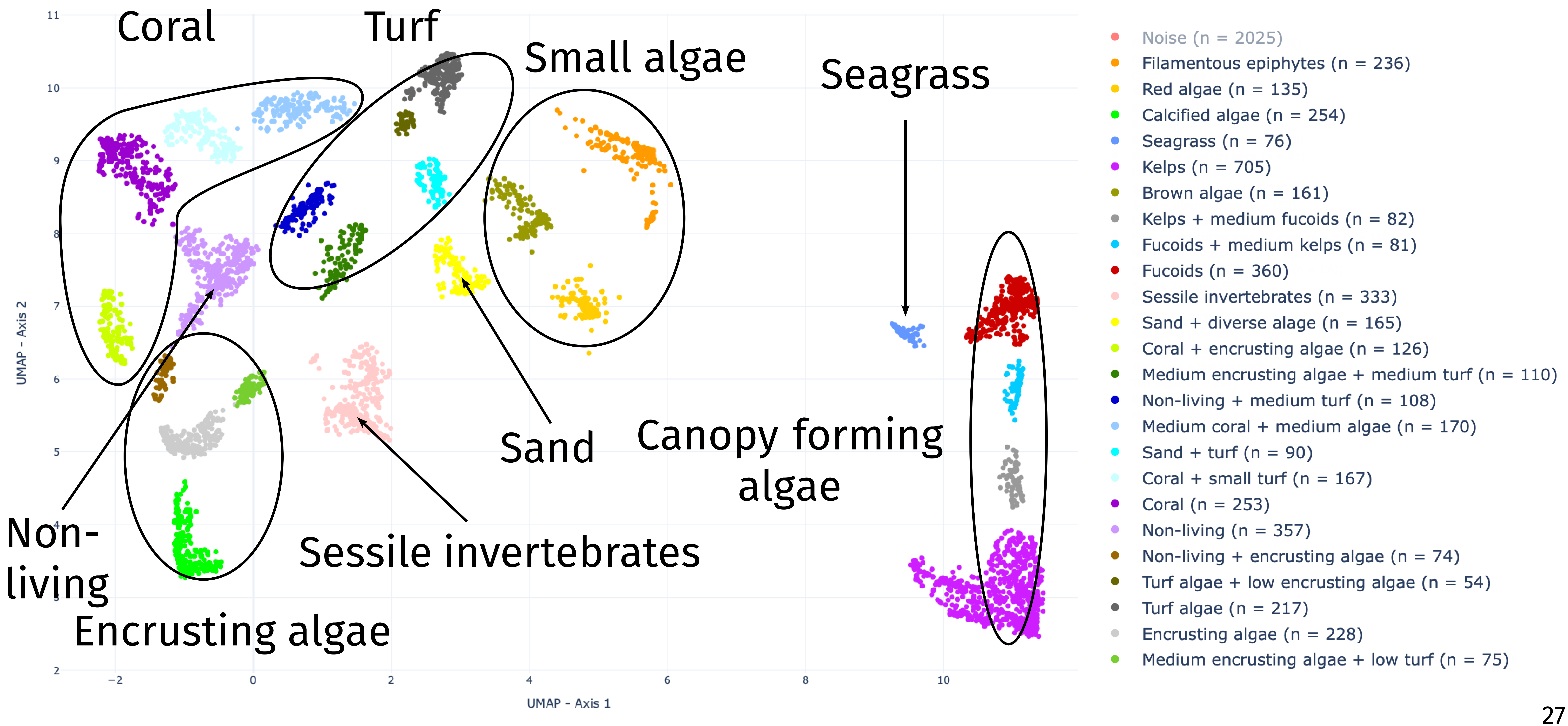
Relative contribution of each variable to this cluster



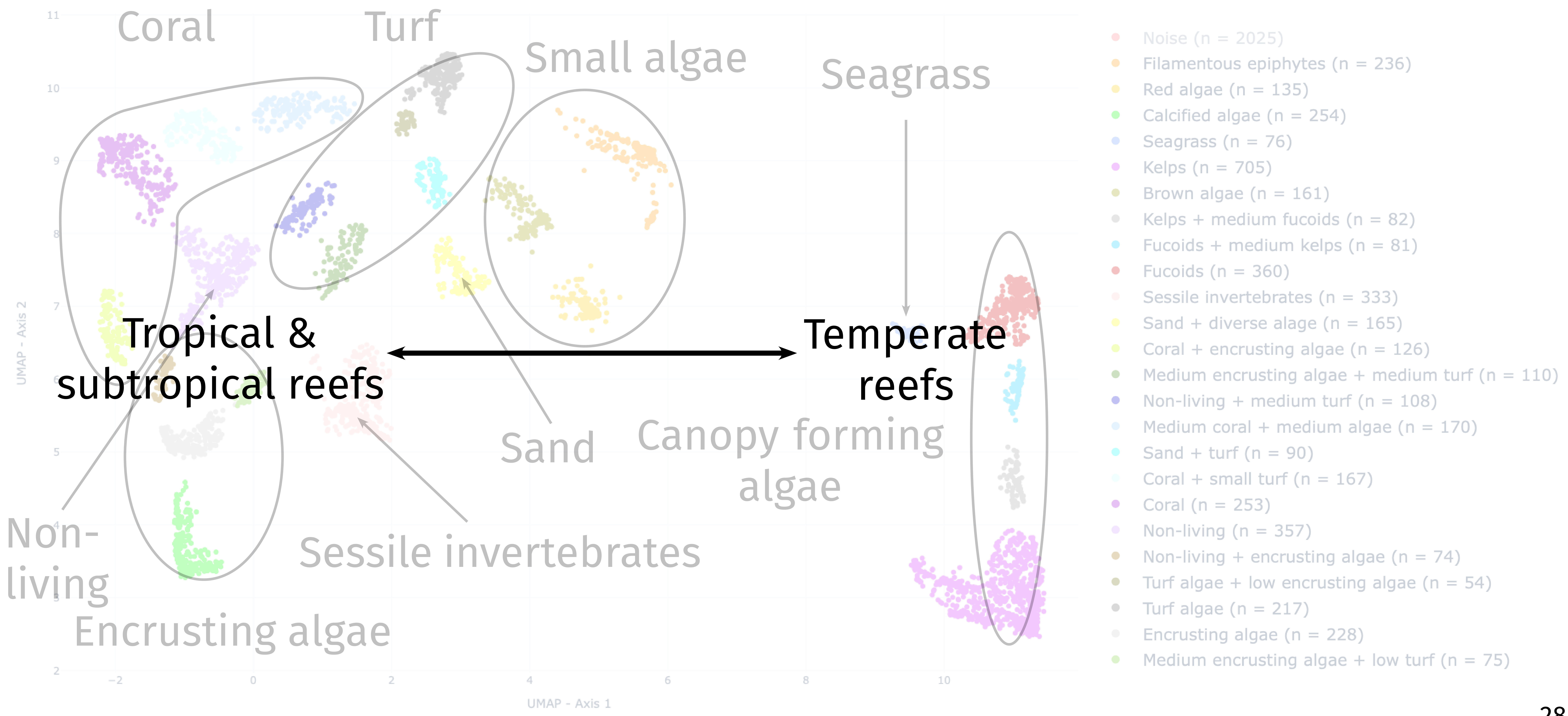
Interpreting overall results & identity of 23 clusters



Interpreting overall results & identity of 23 clusters

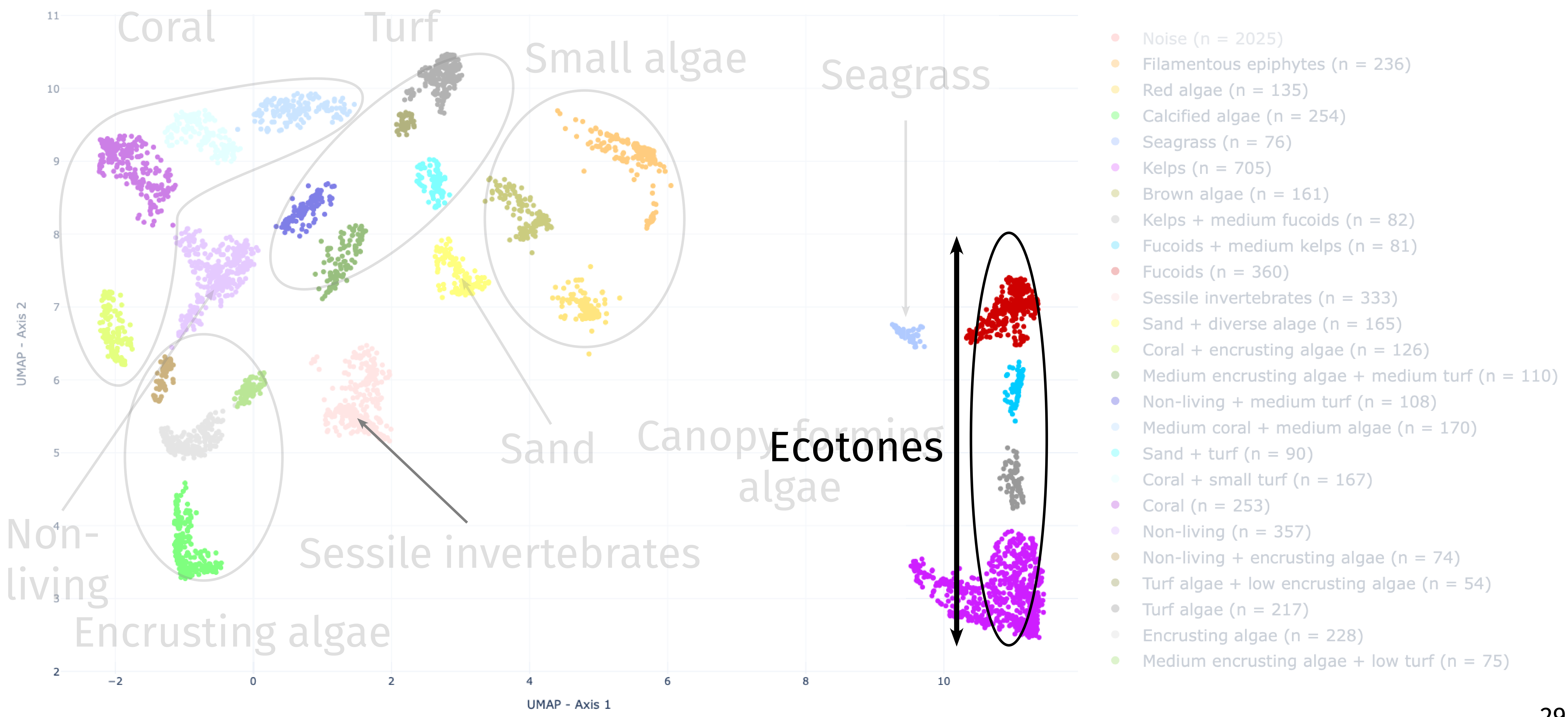


Interpreting overall results & identity of 23 clusters



Tropical & subtropical reefs ↔ Temperate reefs

Interpreting overall results & identity of 23 clusters



4 examples of temperate ecotones identified

Kelp forest

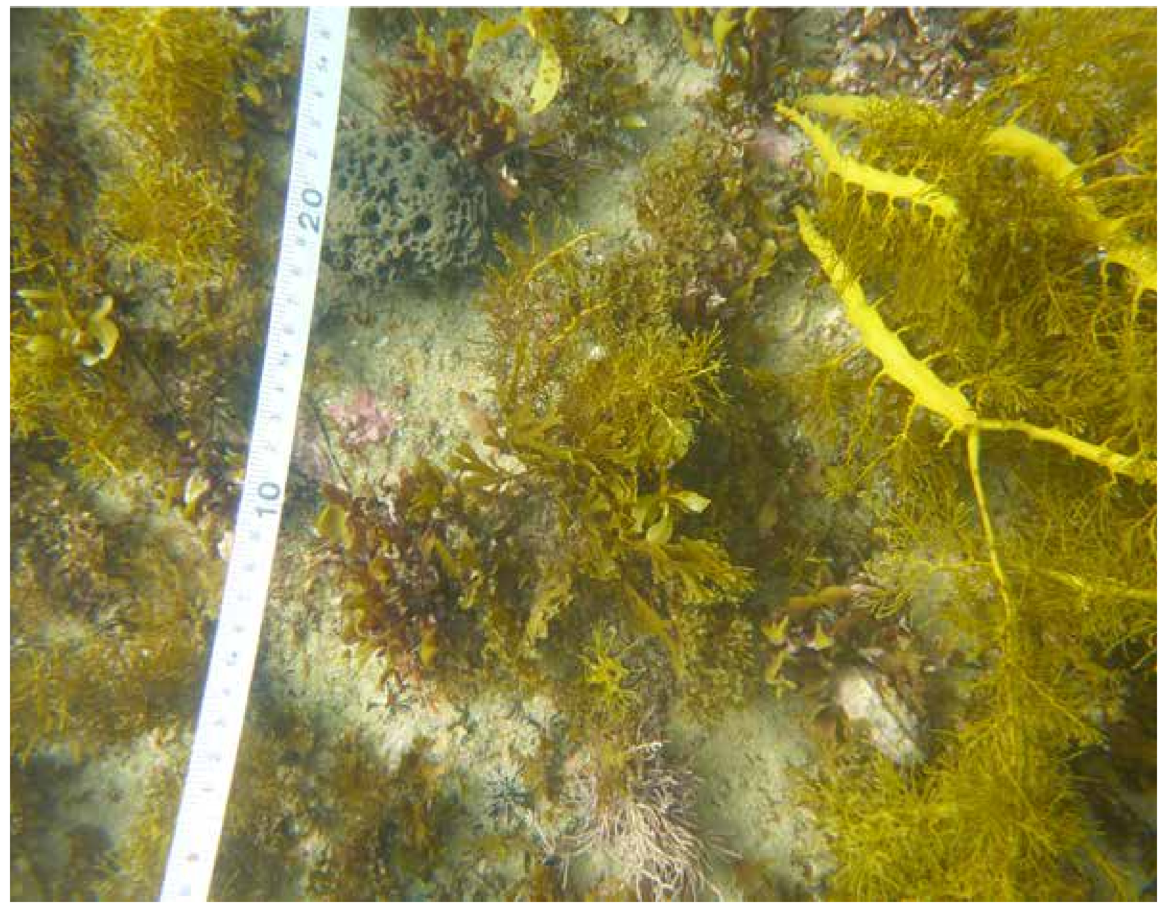
N = 705

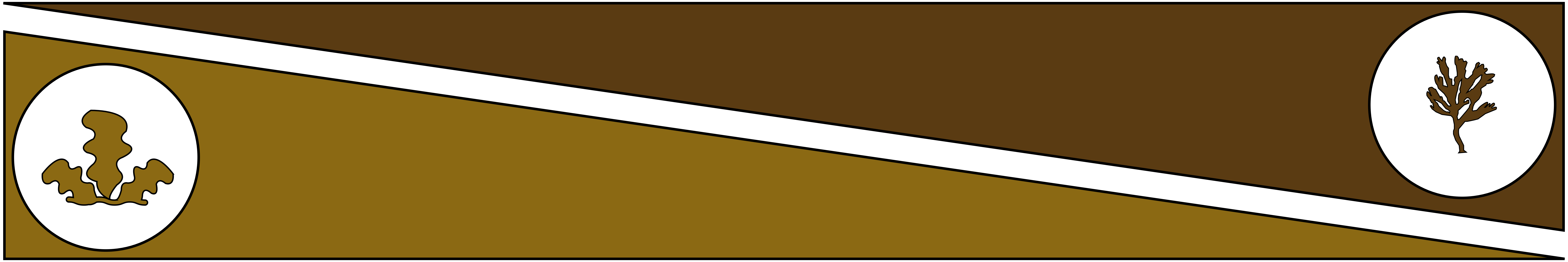
N = 82

N = 81

Fucoid forest

N = 360





Conclusion

Conclusion

1. Development of an effective interpretable clustering pipeline

- Robust against noise
- Scalable to large datasets

©Picture by John Turnbul

Conclusion

1. **Development of an effective interpretable clustering pipeline**
    - Robust against noise
    - Scalable to large datasets
2. **Succesful identification of 23 meaningful clusters based on RLS transects**
    - Large biogeographical patterns across latitudinal gradients
    - Alternative dominant habitat types and ecotones within bioregions

©Picture by John Turnbull

Conclusion

1. **Development of an effective interpretable clustering pipeline**
   - Robust against noise
   - Scalable to large datasets
2. **Succesful identification of 23 meaningful clusters based on RLS transects**
   - Large biogeographical patterns across latitudinal gradients
   - Alternative dominant habitat types and ecotones within bioregions
3. **Work-in-progress:**
   - Quantifying drivers of dominant habitat types occurrence
   - Inclusion of other RLS datasets to define ecological states

©Picture by John Turnbull

Thank you for listening!

✉ clement.violet@ifremer.fr

@ClementVIOLET

@ClementVIOLET@ecoevo.social

clementviolet

Dr Martin
Marzloff

Dr Aurélien
Boyé

Assoc Prof
Rick
Stuart-Smith

Prof Graham
Edgar

References

Bernal-Ibáñez, *et al*. 2021. "The Role of Sea-Urchins in Marine Forests from Azores, Webbnesia, and Cabo Verde: Human Pressures, Climate-Change Effects and Restoration Opportunities." *Frontiers in Marine Science* 8.

Lundberg, Scott M, and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." In *Advances in Neural Information Processing Systems* 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, 4765–74. Curran Associates, Inc.

McInnes, Leland, John Healy, and Steve Astels. 2017. "Hdbscan: Hierarchical Density Based Clustering." *The Journal of Open Source Software* 2 (11).

McInnes, Leland, John Healy, and James Melville. 2020. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *arXiv*.