Integrating Machine Learning with a Genetic Algorithm for Materials Exploration

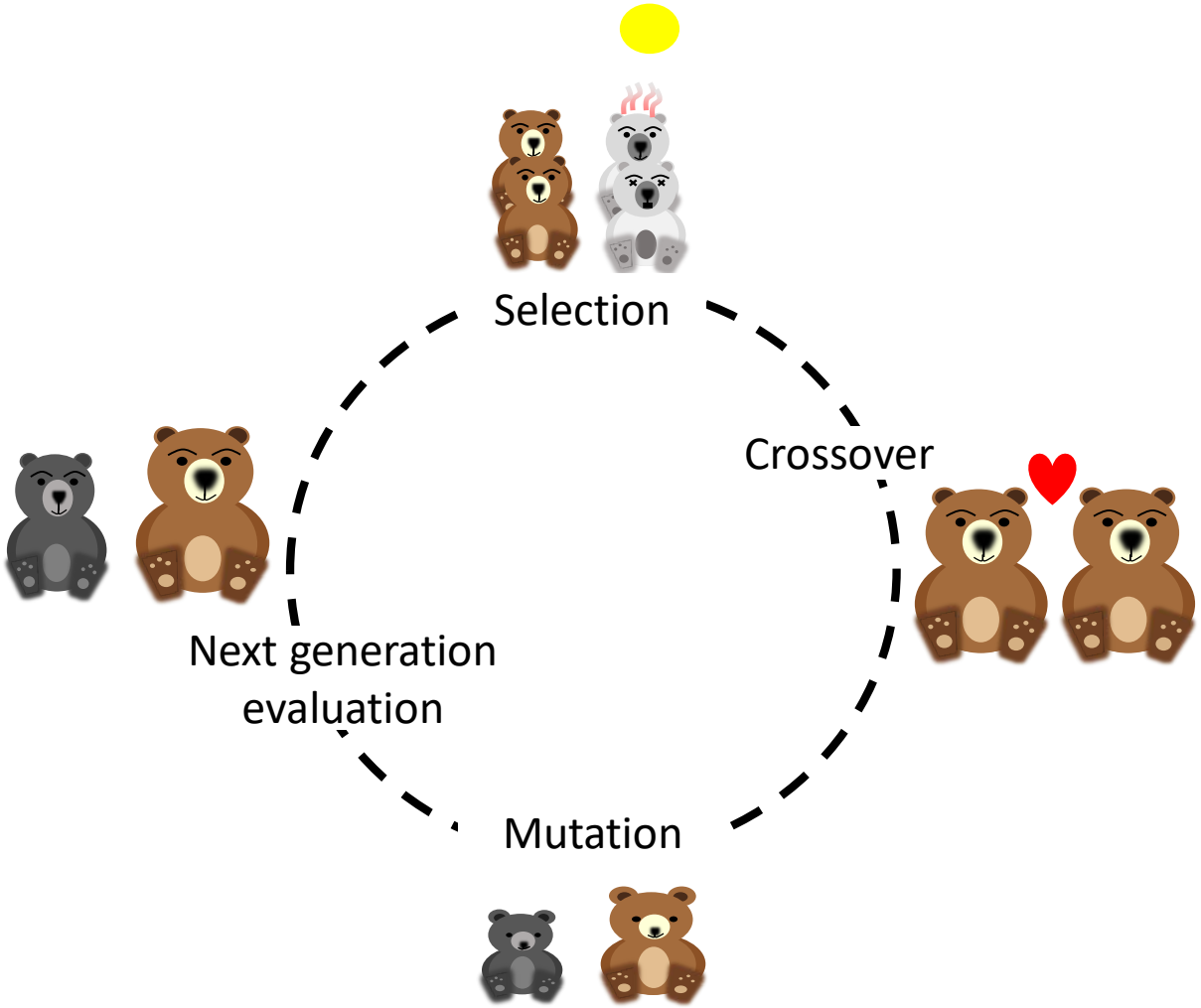
Joseph Kern

J. Kern, L. Chen, C. Kim, and R. Ramprasad, “Design of polymers for energy storage capacitors using machine learning and evolutionary algorithms,” *J Mater Sci*, Sep. 2021, doi: 10.1007/s10853-021-06520-x.

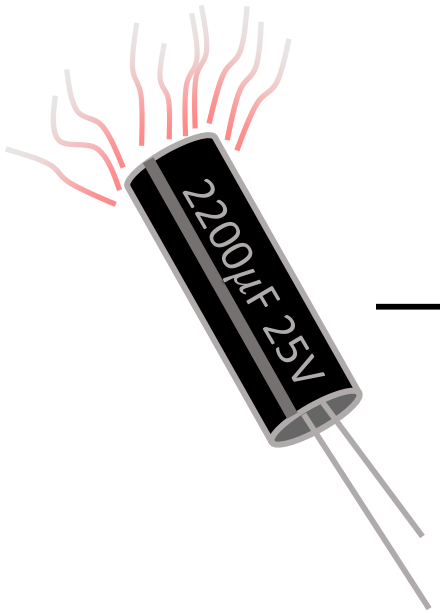
Agenda

- What genetic algorithms are
- What polyga is
- What you need to design a genetic algorithm
- Fitness functions
- Running polyga

The genetic algorithm is inspired by nature.

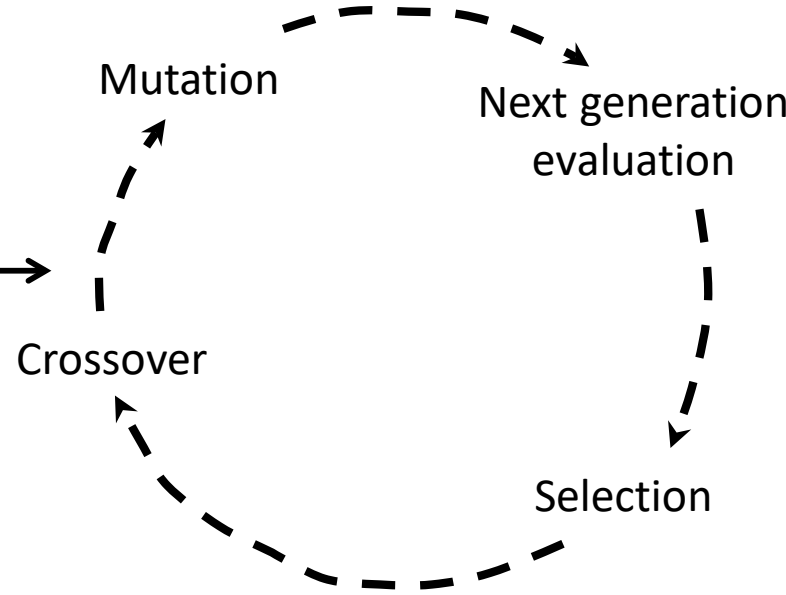


For materials design, we use it to solve the inverse design problem.



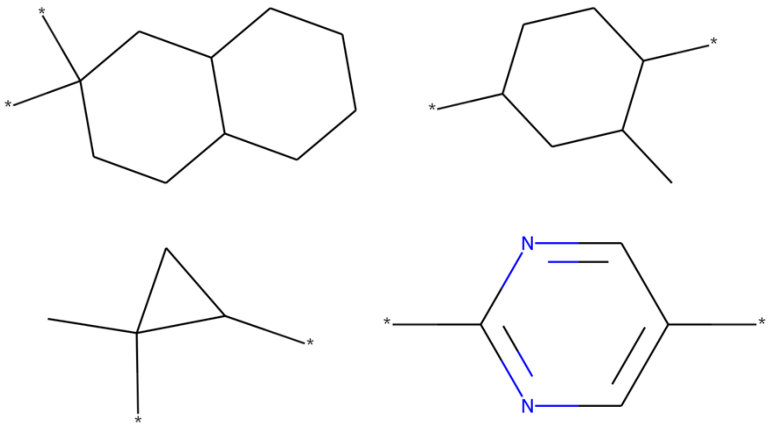
Targets:

- Glass transition temperature (T_g) > 500 K,
- Dielectric Constant (Dk) (100 Hz) > 4,
- Band gap (E_g) > 5 eV,
- Charge Injection Barrier (EIB) > 3 eV,
- Cohesive Energy Density (COE) < 80 $\frac{cal}{cm^3}$

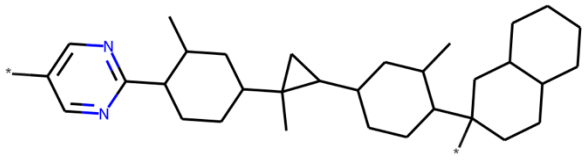


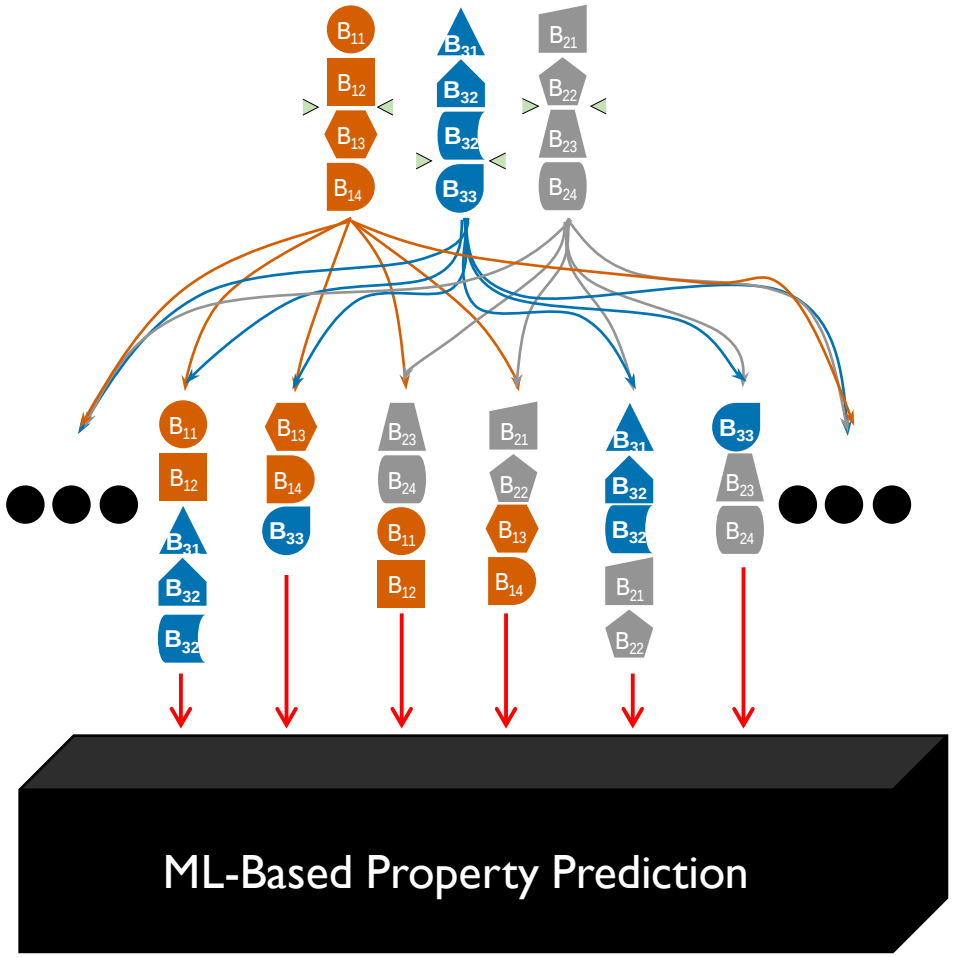
In polyga, we split parents at the center and recombine them

DNA (~3300 blocks)



Polymer

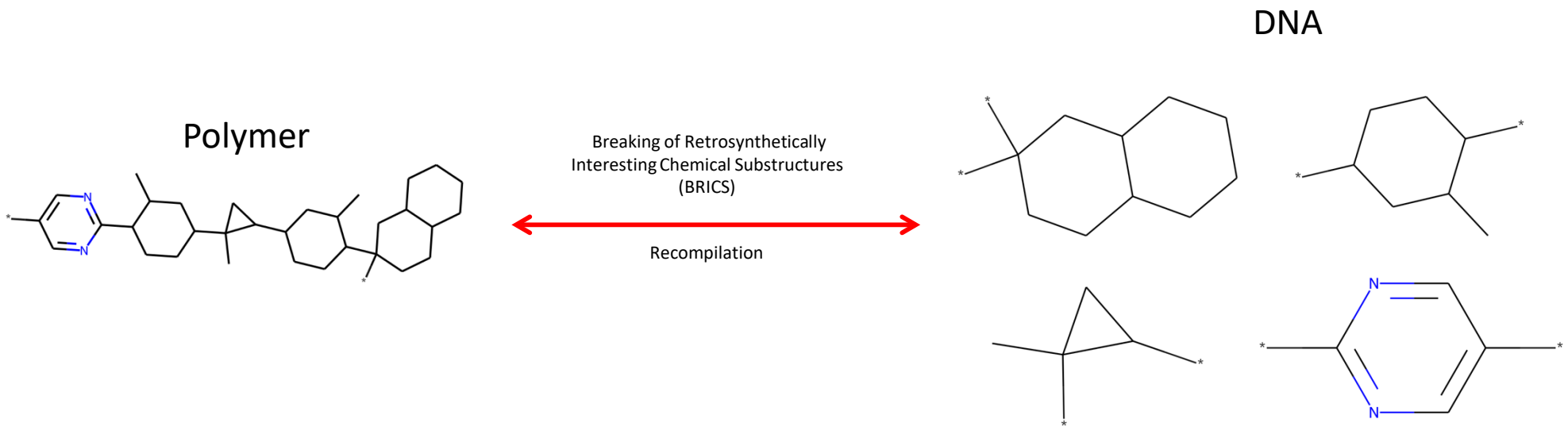




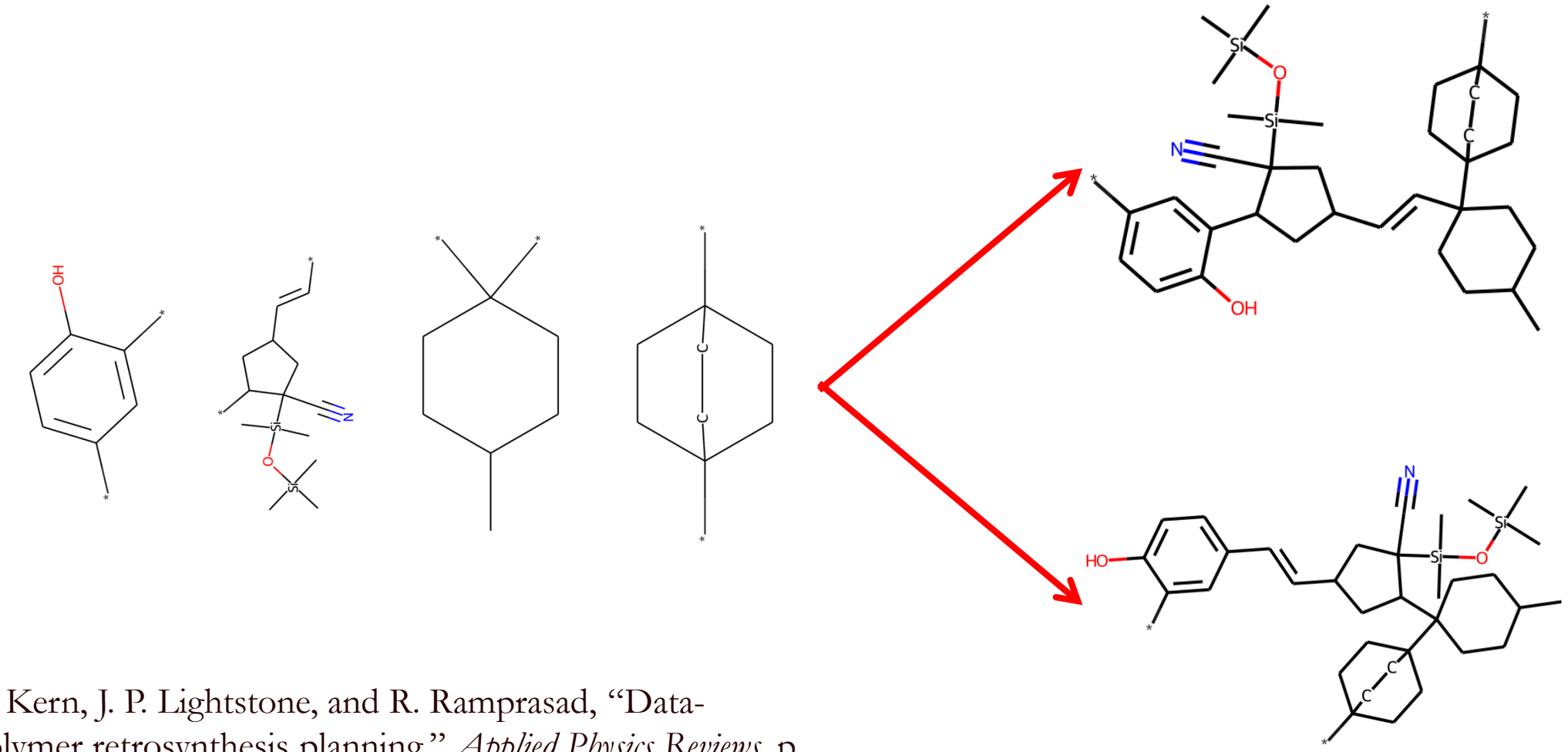
For materials design, we need three things before we can run a genetic algorithm…

1. DNA
2. DNA compilation algorithm
3. Property predictors
    1. Features

I won’t go into the full details of DNA collection because it’s domain specific.

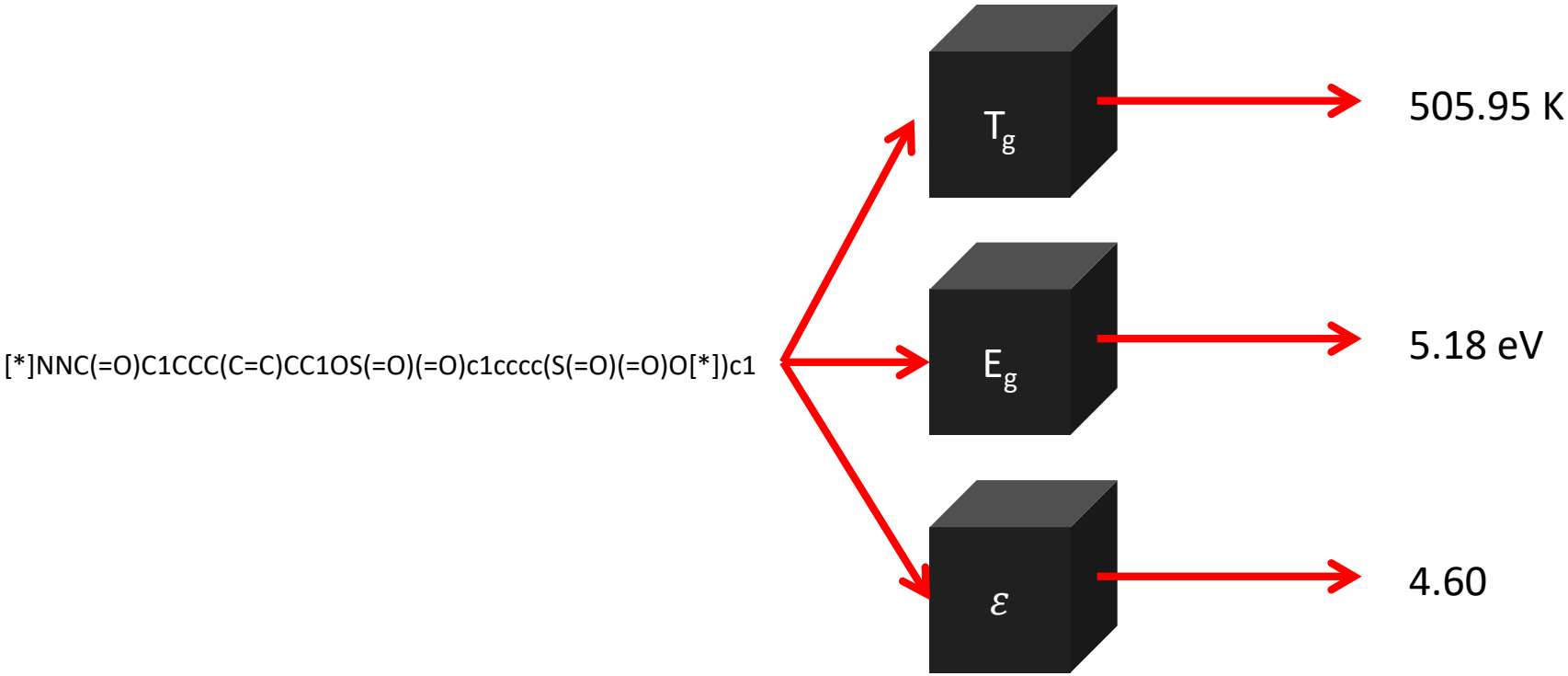


The DNA compilation algorithm is also something you need to design.



L. Chen, J. Kern, J. P. Lightstone, and R. Ramprasad, "Data-assisted polymer retrosynthesis planning," *Applied Physics Reviews*, p. 25, 2021. doi: 10.1063/5.0052962

Property predictors will be models you use to predict the properties of new materials.



Once we have those three things, we can program some criteria for influencing evolution.

Feature	Property
High dielectric breakdown field strength	Large bandgap Large electron charge injection barrier Low cohesive energy density
High thermal stability	High glass transition temperature
Large energy density per volume	Large dielectric constant

Targets:

- Glass transition temperature (T_g) > 500 K,
- Dielectric Constant (Dk) (100 Hz) > 4,
- Band gap (E_g) > 5 eV,
- Charge Injection Barrier (Φ_e) > 3 eV,
- Cohesive Energy Density (e_{coh}) < 80 $\frac{\text{cal}}{\text{cm}^3}$

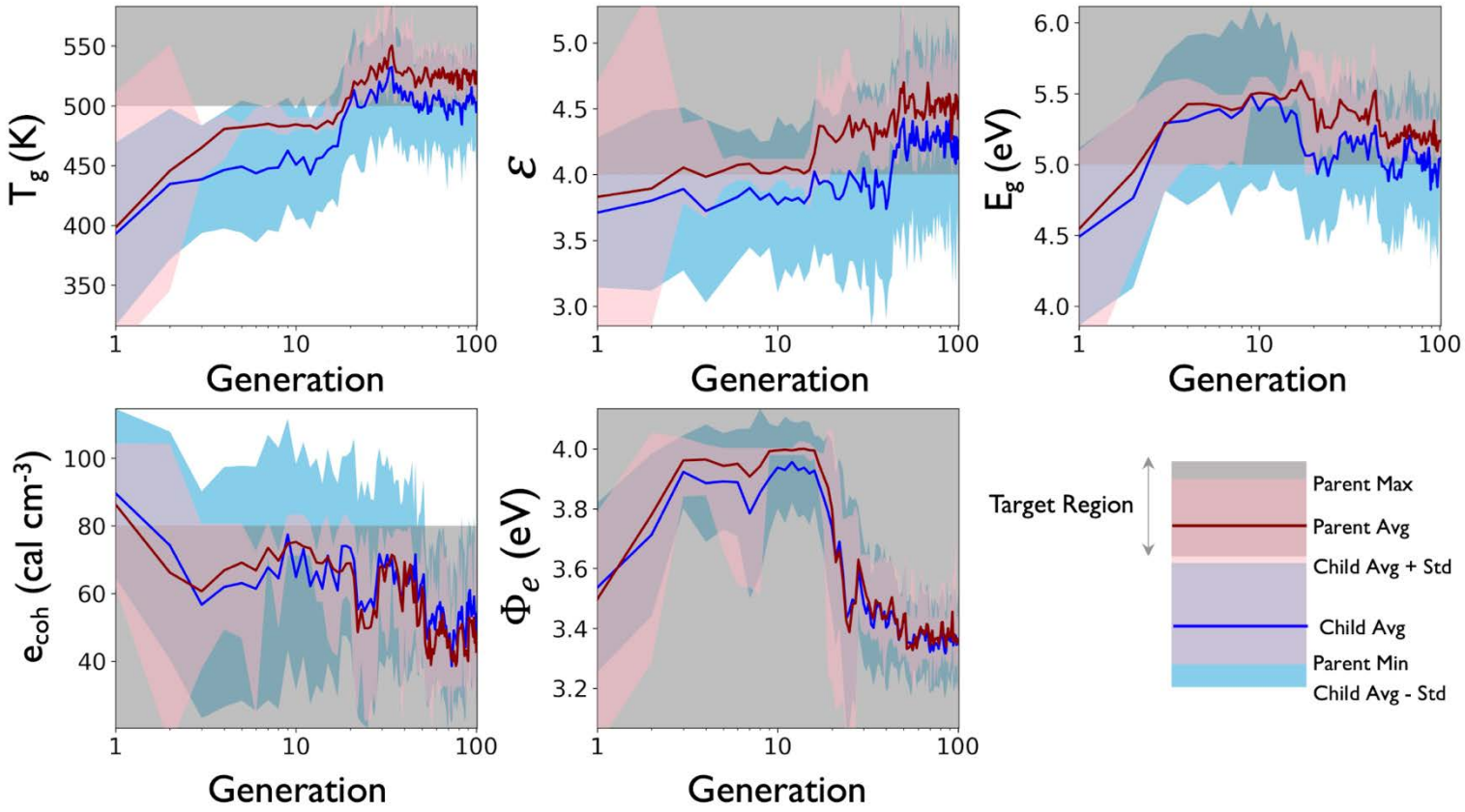
Successful!

Clamping fitness function

$$y_{ij} = \begin{cases} -\max(x_{ij}, target_j), & goal_j = less\ than\ target_j \\ \min(x_{ij}, target_j), & goal_j = greater\ than\ target_j \end{cases}$$

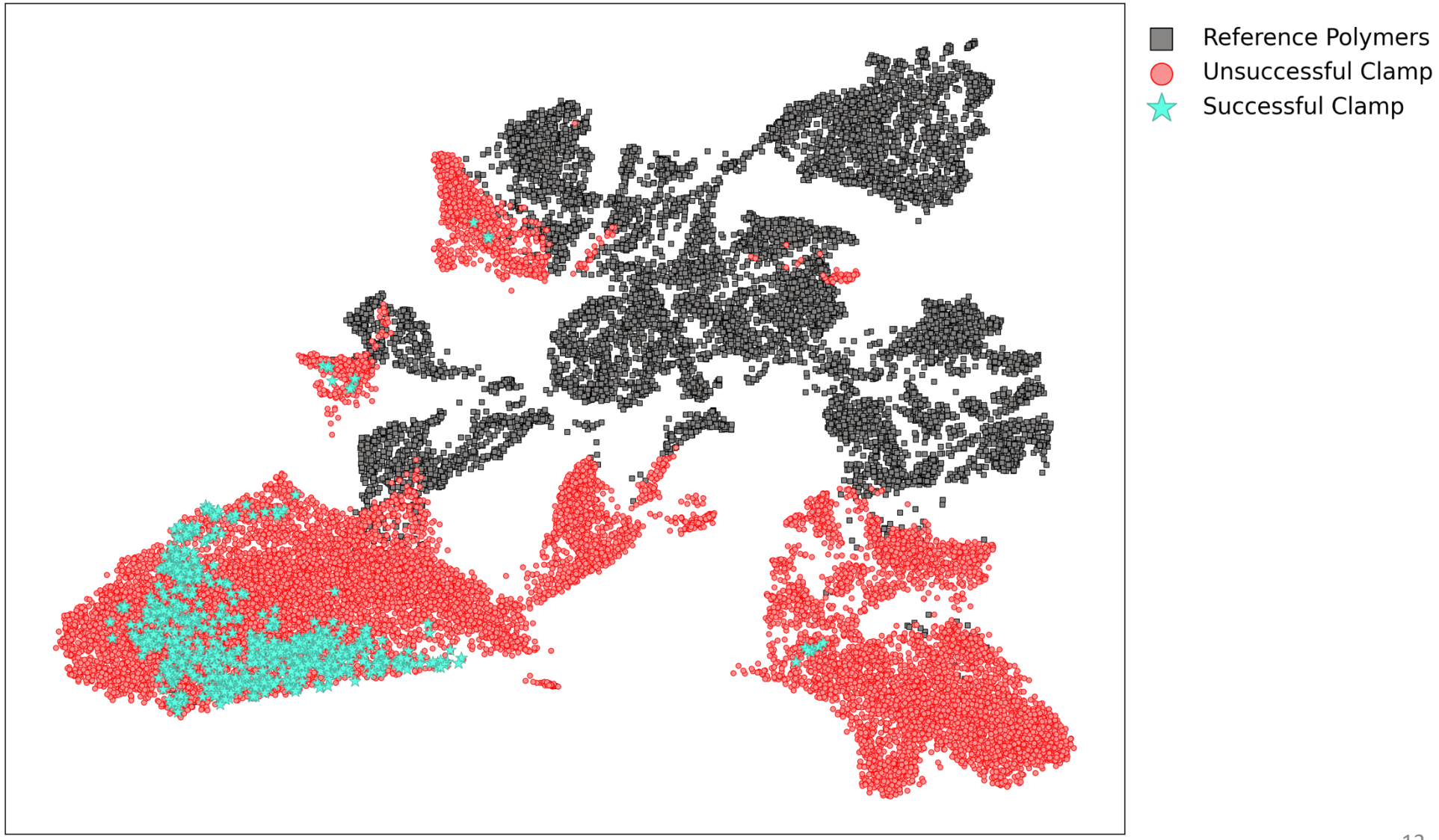
Fitness Score = $\sum minmax(y_{ij})$

Using a clamping fitness function, averages for target properties get closer to their goals with each generation.




```
score = 0.2*tg + 0.2*eg + 0.2*ced + 0.2*eib + 0.2*dk
```

A single run can find many successfully polymers, but diversity is usually limited.



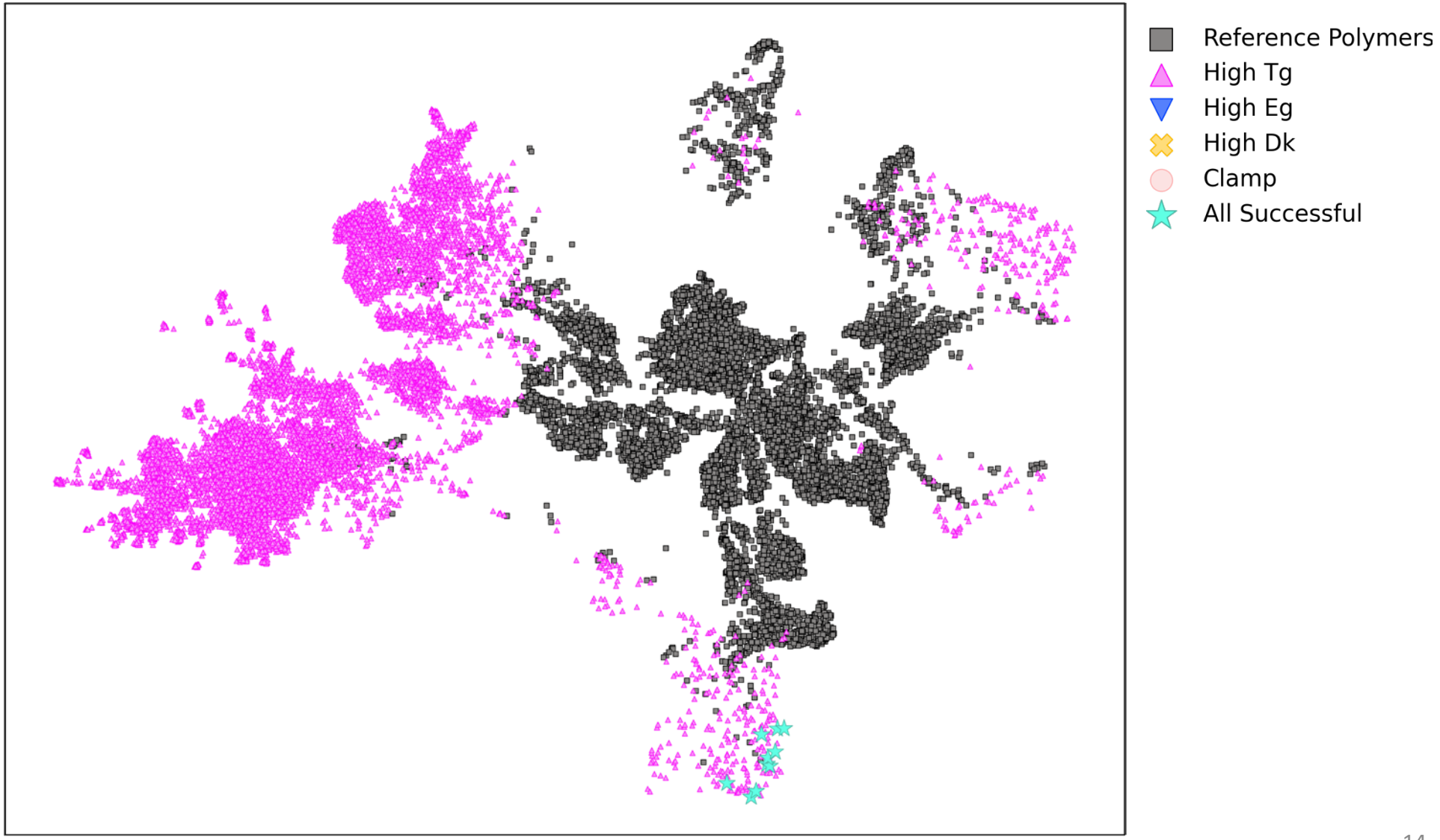
We can create multiple islands with different targets and have migration occur between them to try and achieve increase diversity.

Island Fitness Functions:

1. Maximize T_g
2. Maximize E_g
3. Maximize ε
4. MinMaxScaled linear combination of all properties (Clamp)

J. Verhellen and J. Van den Abeele, “Illuminating elite patches of chemical space,” *Chem. Sci.*, vol. 11, no. 42, pp. 11485–11491, 2020, doi: 10.1039/D0SC03544K.

Using migration, the diversity of the chemical space explored was increased, but the number of successful polymers was not.

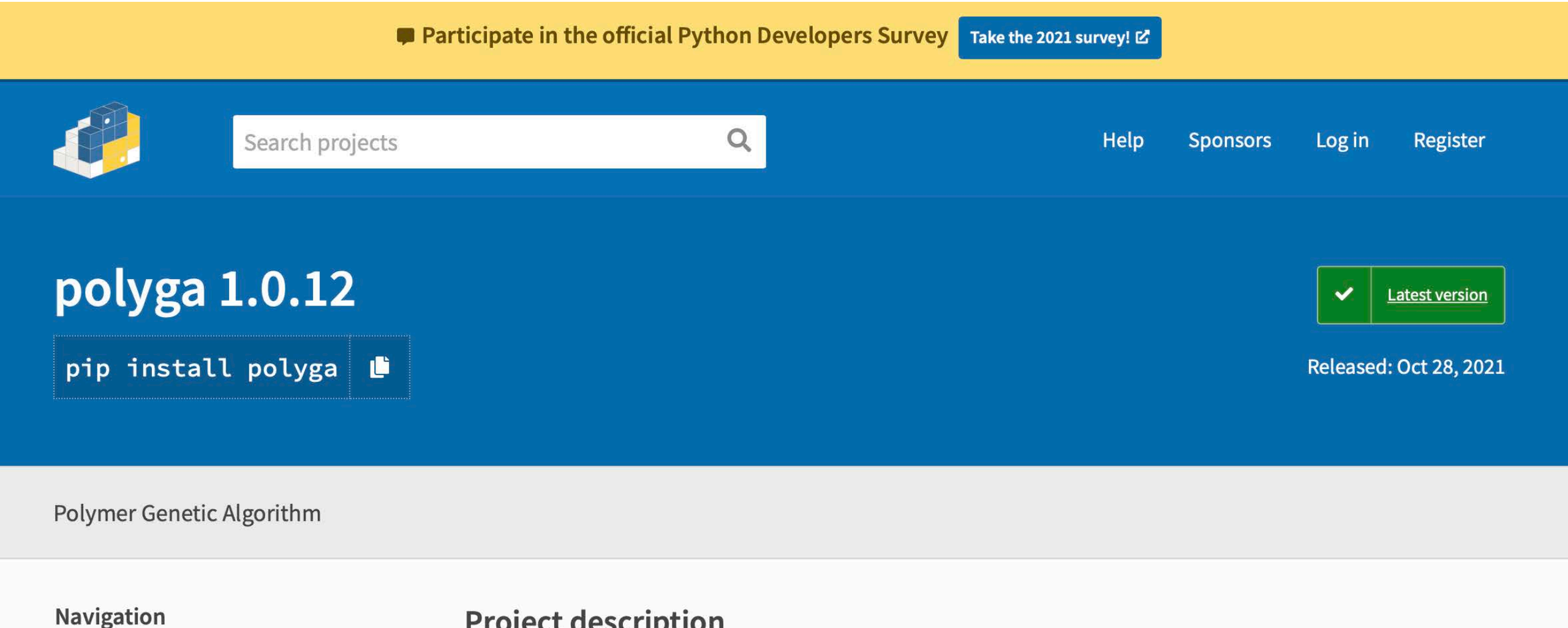


At this point, we can run polyga…

- Crossover
- Mutation
  - Mutation rate
  - Selected randomly from DNA
- Selection
  - Based on fitness functions
  - Based on user desire

https://github.com/Ramprasad-Group/polyga

polyga is pip installable!



https://nanohub.org/resources/polyga