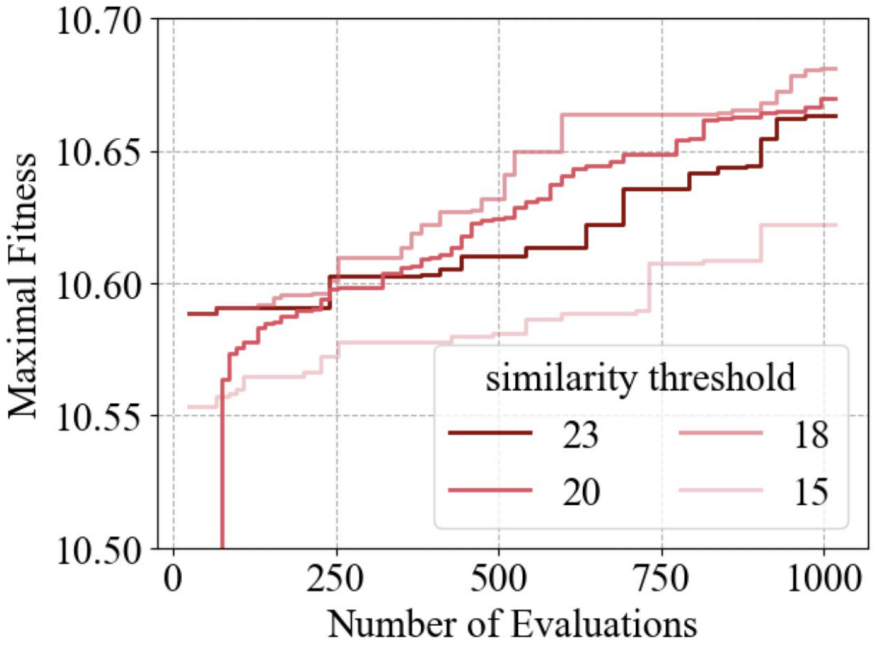
Supplementary material 11:
Further discussion on the similarity threshold

In our diversity reflection mechanism (DiRect), a similarity threshold is needed to decide whether a newly generated robot design is overly similar to existing ones and therefore should undergo modifications by diversity reflection. This threshold is indeed a crucial hyper-parameter that controls the performance of LASeR. The choice of this threshold reflects the extent of diversity that one expects to see in the evolved solutions, and hence should be driven by the user's specific preferences. For instance, setting it as 20 (as we did in our experiments) means that if a newly generated design shares more than 20 identical voxels with any existing solution, it will be modified by DiRect to introduce more variability.

Here, we present some general principles of choosing this parameter. These principles are supported by our additional experiments with several different values of threshold (as shown in **Supplementary figure 11**). High similarity thresholds, like threshold=23, are generally not recommended, as they would hinder the beneficial exploration enabled by LLMs. Conversely, excessively low thresholds (such as threshold=15) might increase diversity but also risk overly aggressive exploration that compromises functionality and, in turn, harms optimization efficiency. We believe this is due to the poor extrapolation performance of LLMs when required to propose robot designs that are much different from given examples. Any moderate values in the middle should give rise to desirable performances. In fact, our findings suggest that a threshold of 18 leads to further performance gains beyond 20, which we have chosen in our study. However, we note that lower thresholds also more frequently trigger DiRect, which means more LLM API calls. Hence, the threshold choice also involves a trade-off between evolutionary performance (including both optimization efficiency and diversity) and computational costs, and should be considered case by case. We believe adaptive threshold scheduling, based on problem specifics and evolutionary outcomes, could be a promising direction for future research.



Supplementary figure 11. Additional experiments with several different similarity thresholds.