Further details on metric calculations

1. City accumulation metric (CAM). One of the strengths of the City Nature Challenge dataset is replication across multiple cities in different geographic regions and environmental contexts. To capture the extent to which urban biodiversity patterns were persistent across cities, we created this metric to evaluate whether species were found in more cities for the highest levels of urbanization than they were in the more natural areas. We created lists of species for each land cover type for each city, then quantified how many cities a species appeared on these lists for each of the land cover types. With a possible 14 cities, each species could have a value of 0-14 for each land cover type. We would expect that species that favored higher levels of urbanization would show up on the highest urbanization lists for more cities than they would for less urbanized lists. As an indicator that higher levels of urbanization intensity lead to greater level of urban homogenization, we would expect that when looking at lists of land cover by each city that there would be more consistency in the lists between cities for the highest levels of development than for the natural and lower ends of development. In other words, if urban homogenization was in effect, lists should not vary much for each of the taxa for the highly urbanized parts of the 14 cities; however, for natural areas and lower urbanization levels, there would much greater variation between the lists.

2. Averaged ranking metric (ARM). Although the previous metric captures the geographic spread of patterns, it does not differentiate between common and rare species. In fact, even if a species is so common as to be characteristic of a region (such as redwood trees along parts of the Pacific coast), if it is geographically endemic, it would appear less prominently than a rarely observed species, so long as that rarely-observed species was observed at least once across many cities. To account for this, we created a ranking metric to evaluate how relatively common species were (compared to similar taxa) within the land use types and cities that they occurred in.

First we broke up the dataset by major taxa groups: 4 plant (monocots, dicots, ferns, and conifers) and 6 animal (birds, insects, reptiles, amphibians, mammals, and gastropods) to allow for better comparisons between similar taxa. Then, for each of the 5 levels of urban intensification (natural, urban-open space, urban-low intensity, urban-medium intensity, and urban-high intensity) for each of the 14 cities, we generated lists that ranked each species based on its number of observations. Ranking species rather than using absolute number of observation comparisons helps to account for the bias of different levels of collecting effort between cities and between land use types. This resulted in 70 lists (5 land cover types x 14 cities) for each of the 10 taxa groups.

With these lists, we were able to qualitatively examine the primarily observed communities for each land cover type for each city. Because we were interested in understanding how patterns change along the urbanization spectrum, we created a ranking metric for each species for each of the 5 land cover types. This ranking metric averaged the rank of a species for each of the 5 levels of urban intensification across all cities that it appeared in. For example, if the honey bee only occurred in urban high intensity land cover in 3 cities, and was ranked as 2nd most observed in Los Angeles, 5th most observed in Dallas, and 10th most observed in Miami, it's ranking metric value would be 5.67 for urban high intensity land cover. We would then be able to compare this value to the honey bee's averaged rankings for the other land cover types to see if it makes up a larger proportion of the urban high intensity insect community than it does in other land cover types.

Each species received a metric value for each of the 5 levels of urban development. From this, we calculated a linear slope for both the CAM and ARM metrics, which is used in Figure 5. All values for species with over 100 observations can be viewed in Supplementary Materials Table 1.